

Ово дело је заштићено лиценцом Креативне заједнице Ауторство – некомерцијално – без прерада¹.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.



¹ Опис лиценци Креативне заједнице доступан је на адреси creativecommons.org.rs/?page_id=74.

"Сва права задржава издавач. Забрањена је свака употреба или трансформација електронског документа осим оних који су експлицитно дозвољени Creative Commons лиценцом која је наведена на почетку публикације."

"Sva prava zadržava izdavač. Zabranjena je svaka upotreba ili transformacija elektronskog dokumenta osim onih koji su eksplicitno dozvoljeni Creative Commons licencom koja je navedena na početku publikacije."



Универзитет у Новом Саду
Природно-математички факултет
Департман за биологију и екологију



МЕТОДЕ У КВАНТИФИКОВАЊУ ФЕНОТИПСКОГ И ГЕНЕТИЧКОГ ДИВЕРЗИТЕТА

др Јасмина Лудошки

Нови Сад, 2023

Назив: Методе у квантификавању фенотипског и генетичког диверзитета

Ауторка: др Јасмина Лудошки, ванредна професорица
Природно-математички факултет, Универзитет у Новом Саду

Рецензенти: др Весна Миланков, редовна професорица
Природно-математички факултет, Универзитет у Новом Саду
др Ана Ивановић, редовна професорица
Биолошки факултет, Универзитет у Београду
др Михаило Јелић, ванредни професор
Биолошки факултет, Универзитет у Београду
др Љубинка Француски Марчетић, виша научна сарадница
Protix BV Холандија

Издавач: Природно-математички факултет, Универзитет у Новом Саду

За издавача: Проф. др Милица Павков Хрвојевић, декан
Природно-математички факултет, Универзитет у Новом Саду

ИСБН: 978-86-7031-630-0

Одлуком Наставно-научног већа Природно-математичког факултета Универзитета у Новом Саду (решење број 0602-07-489/22-9 од 18.01.2023. године) одобрено је штампање и употреба овог рукописа као универзитетског уџбеника.

CIP - Каталогизација у публикацији
Библиотеке Матице српске, Нови Сад

574.1(075.8)

ЛУДОШКИ, Јасмина, 1975-

Методе у квантификавању фенотипског и генетичког диверзитета [Elektronski izvor] /
Јасмина Лудошки. - Нови Сад : Природно-математички факултет, 2023

Начин приступа (URL):

http://www.pmf.uns.ac.rs/studije/epublikacije/biologija/ludoski_kvantifikovanje_fenotipskog_genetickog_diverziteta.pdf. - Opis zasnovan na stanju na dan 23.01.2023. - Nasl. sa naslovnog ekrana. - Bibliografija.

ISBN 978-86-7031-630-0

а) Биодиверзитет -- Квантификација

COBISS.SR-ID 85882377

Садржај

Предговор	v
УВОД	1
I КВАНТИФИКОВАЊЕ МОРФОЛОШКОГ ДИВЕРЗИТЕТА	4
1. Морфометрија	5
1.1. Фенотип и фенотипска варијабилност	5
1.2. Приступи у квантификовању морфолошке варијабилности	9
2. Морфометријски подаци	11
3. Линеарна морфометрија	25
4. Геометријска морфометрија	29
4.1. Суперимпозиција	30
4.2. Простор облика	35
4.3. Визуелизација варијабилности облика	39
5. Статистичка анализа морфометријских података	44
5.1. Тестирање хипотеза	45
5.2. Униваријантне статистичке методе	50
5.3. Мултиваријантне статистичке методе	57
6. Анализа морфолошког диверзитета - примери	79
6.1. Просторна дистрибуција морфолошког диверзитета врсте <i>Telestes dabar</i> (fam. Leuciscidae)	80
6.2. Интерспецијска и интраспецијска варијабилност врста <i>Ochlerotatus caspius</i> и <i>O. dorsalis</i> (Diptera, Culicidae)	85
II КВАНТИФИКОВАЊЕ ГЕНЕТИЧКОГ ДИВЕРЗИТЕТА	89
7. Предеона генетика	90
8. Молекуларни подаци	92
8.1. Молекуларни маркери	95
8.2. Типови молекуларних података	100

9. Мерење генетичке варијабилности	100
10. Процена генетичке структуре популације и детектовање баријера	102
10.1. Параметри генетичке структуре популације	103
10.2. Методе за идентификовање просторних образаца дистрибуције генетичког диверзитета	109
10.2.1. Анализа главних компоненти	111
10.2.2. Методе Бајесове кластер анализе	112
10.2.3. Методе просторне аутокорељације	120
10.3. Визуелизација просторних образаца дистрибуције генетичког диверзитета: метода интерполације	121
10.4. Детектовање потенцијалних баријера за проток гена	125
10.4.1. Детектовање баријера у просторној Бајесовој кластер анализи	125
10.4.2. Детектовање баријера методама детекције ивица	127
11. Корелисаност генетичког, географског и фенотипског дисконтинуитета	131
ПРИЛОГ	135
Прилог 1. Корак по корак упутство за програме <i>tps</i> серије: <i>tpsUtil</i>	136
Корак по корак упутство за програме <i>tps</i> серије: <i>tpsDig</i>	139
Прилог 2. Корак по корак упутство за програме <i>tps</i> серије: <i>tpsRelw</i>	146
Прилог 3. Корак по корак упутство за програме <i>tps</i> серије: <i>tpsRegr</i>	150
Прилог 4. Корак по корак упутство за програм <i>MorphoJ</i>	154
ЛИТЕРАТУРА	161

Предговор

Уџбеник **Методе у квантификовању фенотипског и генетичког диверзитета** намењен је првенствено студентима Основних академских студија биологије и екологије на Департману за биологију и екологију Природно-математичког факултета Универзитета у Новом Саду који на другој, односно трећој години похађају наставу из изборних предмета Геометријска морфометрија у биолошким истраживањима и Методе у квантификовању биолошког диверзитета. Садржај уџбеника је написан тако да у потпуности покрива наставни план наведених предмета, али може се користити и као додатна литература за поједине наставне јединице на предметима Еволуциона биологија и Основе еволуционе биологије. Такође, уџбеник је намењен и свим члановима академске заједнице заинтересованим за сагледавање и разумевање биолошке разноврсности (= биодиверзитета), његову угроженост и конзервацију.

Настао са циљем да читаоца упозна са теоријским поставкама и практичном применом актуелних метода за квантификовање и анализу дистрибуције фенотипског и генетичког диверзитета, текст уџбеника је подељен у две целине. Први део је посвећен упознавању са морфометријским техникама и методама статистичке анализе података. У другом делу дат је кратак увод у предеону генетику, као и преглед метода за процену генетичке структуре популација и детектовање баријера, те метода за интегрисање морфометријских и генетичких података. Такође, рукопис укључује и прилоге у којима су дата упутства за употребу софтвера за прикупљање и анализу морфометријских података.

Велику захвалност дугујем рецензентима др Весни Миланков, редовној професорици Природно-математичког факултета Универзитета у Новом Саду, др Ани Ивановић, редовној професорици Биолошког факултета Универзитета у Београду, др Михаилу Јелићу, ванредном професору Биолошког факултета Универзитета у Београду и др Љубинки Француски Марчетић, вишој научној сарадници на врло корисним сугестијама и примедбама које су значајно допринеле коначном садржају и изгледу уџбеника. Захвална сам и докторандкињи Милицы Лукач на уступљеним подацима и сликама, као и корекцијама и коментарима приликом читања прве верзије текста.

УВОД

Према Конвенцији о биолошкој разноврсности (*Convention on Biological Diversity*; CBD) биолошка разноврсност (= биолошки диверзитет, биодиверзитет) се дефинише као свеукупна различитост живих организама из свих извора, укључујући копнене, морске и друге водене екосистеме и еколошке комплексе чији су они део; то укључује диверзитет унутар врсте, између врста и екосистема (CBD, 2001). Осим што обухвата све организме, врсте и популације, генетичку варијабилност између њих, све сложене скупове заједница и екосистема које они чине, биодиверзитет се односи и на међусобну повезаност гена, врста и екосистема и њихове интеракције са животном средином.

Биодиверзитет, чинићи сложену мрежу живих бића која интерагују међусобно унутар екосистема, региона или глобално, покреће функционисање екосистема кроз небројене узајамне интеракције са физичким и хемијским компонентама животне средине. Значи, биодиверзитет има кључну улогу у функционисању екосистема у променљивој средини, а тиме и у одржавању живота на земљи. Бројне су функције које има, укључујући (Royal Society, 2003):

- одржавање равнотеже у екосистемима - свака врста има специфичну улогу у екосистему. Организми производе и складиште хранљиве материје, разлажу органску материју, одржавају стабилност климе, учествују у заштити и пречишћавању водних ресурса, формирању и заштити земљишта и слично;
- обезбеђивање биолошких ресурса - биодиверзитет садржи обиље ресурса за исхрану (сва храна се директно или индиректно добија од биљака и других фотосинтетских организама), производњу лекова и других фармацеутских производа, те употребу у разне друге сврхе (гориво, влакна, гума, смоле и слично);
- естетски и културни значај - врсте су често саставни део верских, културних и националних идентитета људских заједница. Бројне су и друге друштвене користи које људи имају од биодиверзитета, од образовања и истраживања (пружа информације које нам помажу да разумемо свет природе и његово порекло), преко туризма и рекреације (рекреативне активности посматрања птица, планинарења, камповања и риболова базирају се на биодиверзитету), до инспирације за уметнике.

Упркос суштинском значају биодиверзитета за опстанак живота на земљи (укључујућу и опстанак нас самих), данас смо суочени са губитком биодиверзитета брзином највећом познатом у геолошкој историји Земље, а које је у највећој мери последица антропогеног утицаја. Директан резултат људске активности (неодрживо и неодговорно коришћење ресурса) и непрестаног раста наше популације (у новембру 2022. године људска популација је достигла бројност од 8 милијарди) су: губитак и уништавање природних станишта који воде ка губитку врста, популација и екосистема; промене у саставу екосистема; интродуковање егзотичних (алохтоних) врста што утиче на популације аутохтоних врста и може пореметити читаве екосистеме; прекомерна експлоатација (претерани лов, риболов или сакупљање) што доводи до драстичног смањења бројности популација, па и њиховог нестанка; загађење животне средине;

глобалне климатске промене (Raven, 2002; Royal Society, 2003). Врсте и популације које немају довољно генетичког потенцијала (генетичку варијабилност) да се адаптирају на нове услове или способност да мигрирају у станишта са повољнијим условима могу бити изгубљене. А свака промена, губитак или угрожавање стања биодиверзитета може нарушити равнотежу и опстанак екосистема.

У циљу сагледавања и разумевања биодиверзитета, његове дистрибуције, угрожености и дефинисања програма његове заштите, квантификовање биодиверзитета је од примарне важности. Биодиверзитет се може квантификовати на различитим нивоима биолошке организације - од молекуларног до нивоа екосистема. Хијерархијску организацију биодиверзитета чине генетички ниво, ниво врсте/популације, ниво заједнице/екосистема и ниво предела или региона (Gaines и сар., 1999; Laffont и сар., 2011):

- Генетички диверзитет се односи на ширину генетичке варијабилности унутар и између популација и врста. Обухвата све различите гене присутне у геному сваке појединачне биљке, животиње, гљиве и микроорганизама. Генетички диверзитет је неопходан предуслов за будуће адаптивне промене или еволуцију. Пошто је генетичка варијабилност од суштинског значаја за опстанак популација и врста у променљивим срединама, губитак генетичког диверзитета смањује шансу за њихово преживљавање. Стога је очување генетичког диверзитета популација приоритет у конзервацији и менаџменту врста;
- Специјски диверзитет (= диверзитет врста) се односи на укупан број врста које се могу наћи у одређеном региону. То је биодиверзитет на најосновнијем нивоу. Најчешће се за процену биодиверзитета користи мера богатство таксона (*taxon richness*) која подразумева број врста (или виших таксона) присутних у неком екосистему, региону или одређеном подручју. Диверзитет врста је функција броја присутних врста (богатство таксона) и релативне бројности сваке од њих. Иако су специјски диверзитет и богатство врста често у позитивној корелацији, постоје ситуације у којима је повећање диверзитета врста праћено смањењем богатства врста;
- Еколошки (екосистемски) диверзитет обухвата сва различита станишта, биолошке заједнице (врсте које заузимају одређено подручје и интеракције између њих) и еколошке процесе, као и варијабилност унутар појединачних екосистема. Биолошка заједница са својим физичким окружењем чини екосистемом.

Процена генетичког диверзитета обично се заснива на проучавању морфолошких карактера, алозима и/или ДНК секвенци. Стога ће у наставку овог рукописа фокус бити на методама које се користе за квантификовање морфолошке и генетичке варијабилности.

ДЕО I
КВАНТИФИКОВАЊЕ
МОРФОЛОШКОГ ДИВЕРЗИТЕТА

1. Морфометрија

1.1. Фенотип и фенотипска варијабилност

Опште је познато да је фенотип организма (= скуп детектабилних особина једног организма) резултат експресије његовог генотипа (= генетичка конституција јединке), утицаја фактора животне средине, те интеракције између њих. Фенотип се реализује током процеса индивидуалног развића (онтогенезе) помоћу сложеног скупа фактора и њихових интеракција, укључујући ДНК, РНК, протеине, интеракције на нивоу ћелија и ткива, развојне модуле (компоненте организма у развоју; Schlosser и Wagner, 2004) и различите аспекте биотичке и абиотичке средине (Wolf, 2002). Другим речима, коначни (спољашњи) фенотипови нису регулисани само комбинованим генотиповима, већ зависе и од бројних спољашњих (еколошких) фактора (попут исхране, срединских услова и слично), као и од окружења са којима се сусрећу појединачне ћелије и/или ткива/органи (te Pas и сар., 2017). Средински фактори могу да утичу и на експресију и функционисање скупова гена/протеина/метаболита и/или комуникацију између ћелија/ткива, било директно на нивоу синтезе или стабилности РНК, или путем епигенетичких промена. Епигенетичке промене подразумевају процесе који доводе до хемијске модификације (попут метилације, ацетилације, фосфорилације, убиквитинације и сумоилације) ДНК молекула (не доводећи до промене саме секвенце) и хистона, а што утиче на експресију гена (te Pas и сар., 2017). Тако ће исти гени (генотип) током развића под контролом епигенетичких процеса кроз диференцијалну генску експресију дати различит исход (фенотип) у различитим ткивима/органима/организмима. Такође, и посттранслационе модификације протеина могу утицати на активност протеина, а од чега може зависити и ефикасност транслације, као и проток метаболита кроз развојне путеве.

Развиће организма је веома сложен процес који је у исто време и флексибилан (организам може да одговори на срединске услове променама у развићу, а тиме и резултирајућем фенотипу), али и веома робустан (све развојне промене у различитим деловима су координисане што резултира формирањем високо интегрисаног, функционалног организма). Истраживања су показала да је ова истовремена флексибилност и робусност базирана на модуларној архитектури организама (Klingenberg, 2004). Модуларна организација (= модуларност) је утврђена на многим нивоима организације, од молекуларне структуре, преко појединачних гена, до планова телесне организације организама. Модули су јединице у оквиру развојног система (организма) које су унутар себе тесно повезане вишеструким интеракцијама између својих делова и релативно независне од других таквих јединица, а део су веће јединице (на пример, различити делови тела организама понашају се као модули и показују одређени степен међусобне аутономије у складу са њиховим развојним пореклом и функцијама). С обзиром да ове јединице могу да функционишу у различитим контекстима (развојни, генетички, функционални, еволуциони) и одвојено пролазе кроз

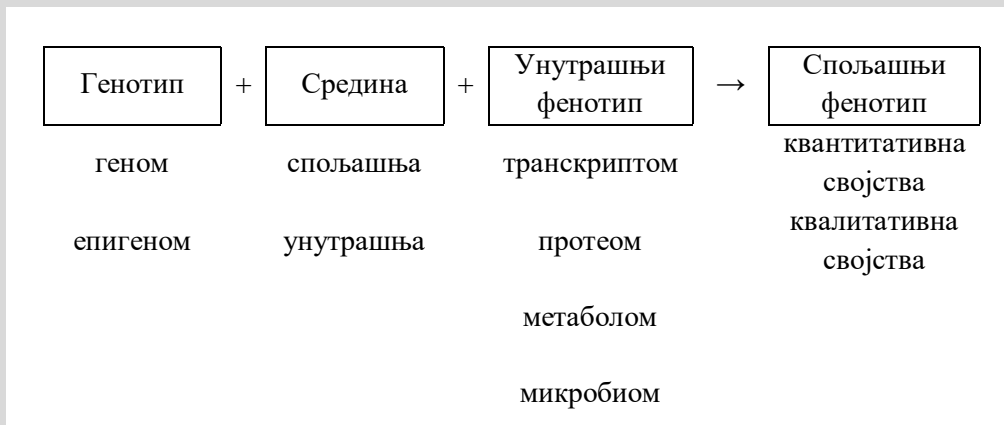
развојне и еволуционе промене, модулarna организација омогућава да се одређени делови тела мењају без ометања функције других делова (Hall, 2003; Klingenberg, 2004).

Модуларност је кључни концепт који повезује генотип и фенотип јер модули обухватају и обједињују различите биолошке нивое (Mabee, 2006). Генетички модули представљени хијерархијски организованим генским мрежама (*regulatory gene networks*) и генским каскадама (*gene cascades*) управљају генском експресијом на нивоу информационе РНК и протеина и одређују функцију ћелије. На овај начин, генетички модули посредују између генотипа и морфогенетских јединица. Морфогенетске јединице (модули) су фундаменталне ћелијске компоненте структурне биолошке хијерархије гена → ћелија → ткива → органа → организама → врста од којих је конструисана морфологија организма (Hall, 2003). То су групе ћелија (на пример, имагинални дискови ларви инсеката, пупољци екстремитета ембриона кичмењака или замеци зуба) од којих се развијају специфичне морфолошке структуре или органи. Образац развића ткива и телесних структура у процесу морфогенезе одређују сигнални молекули, морфогени, тако што стварајући локални градијент концентрације биохемијских сигнала иницирају процес диференцијације неспецијализованих матичних ћелија у различите типове ћелија (а које потом изграђују ткива и органе) (Hall, 2003). Такође, овако модулarna и хијерархијска организација омогућава да групе сличних ћелија примају срединске и епигенетичке сигнале који омогућавају организмима да се развију, адаптирају на своје окружење, модификују свој развој и преведу ефекте мутација у фенотипске промене у развојном и еволуционом оквиру. Дакле, у хијерархији процеса који повезују генотип и фенотип (генотип → генетички модули → морфогенетске јединице → епигенетски процеси → фенотип), епигенетски процеси повезују морфогенетске јединице са фенотипом (Hall, 2003). Контрола морфогенезе је централни елемент у еволуционој развојној биологији (*Evolutionary Developmental Biology; evo-devo*). Модификовањем развојних процеса настају фенотипске промене, а може доћи и до појаве нових карактеристика (еволуционе новине). Наиме, утицаји у раном периоду развића вероватно ће довести до губитка елемената или до великих структурних или функционалних промена, док утицаји у каснијој фази вероватно ће утицати на мање промене (што не значи и мање значајне) у величини и облику структура (Hall, 2003).

Сви фенотипови су резултат биолошких процеса који се могу одвијати на различитим нивоима (на пример, на нивоу експресије гена, РНК, експресије протеина, метаболита, микробиома и слично), те се стога фенотипови могу процењивати/мерити на сваком од тих нивоа. Међутим, ови биолошки нивои нису међусобно независне, самосталне јединице. Бројне повратне спреге унутар и између нивоа обезбеђују регулацију експресије генома, а тиме и сложених особина. Дакле, компоненте активне на одређеном биолошком нивоу, а које доприносе развоју спољашњег (екстерног) или коначног фенотипа саме по себи чине фенотипове. Тако, профили експресије транскриптома, протеома и метаболома представљају унутрашње фенотипове (интерне фенотипове, ендифенотипове; називају се тако јер нису лако видљиви споља) који укључују ћелијски ниво, ткивни ниво и физиолошка својства, и утичу на обликовање спољашњих фенотипова (на пример, морфологију и понашање) (Houle и сар., 2010;

Fontanesi, 2016; Benis и сар., 2017; te Pas и сар., 2017). Такође, и спољашњи фенотипови могу обликовати средину коју јединка заузима, стварајући тако сложене повратне везе између гена, окружења и фенотипова (Houle и сар., 2010). Фенотипови ступају у интеракцију са средином и до изражаја долазе разлике у адаптивној вредности између њих, а диференцијална адаптивна вредност је повод за деловање природне селекције. Судбина генетичке варијанте која се реализује кроз одређени (крајњи) фенотип зависиће од укупног ефекта који она има на адаптивну вредност свих фенотипова (унутрашњег и спољашњег) на које утиче (Houle и сар., 2010).

Однос између фенотипа, генотипа и средине. Генотип и средина утичу на унутрашњи фенотип организма и заједно регулишу спољашњи (коначни) фенотип, односно особину. Извор: модификовано из Fontanesi (2016), Benis и сар. (2017), te Pas и сар. (2017).



Геном - укупна генетичка информација организма.

Епигеном - скуп једињења који модификују или обележавају молекулу ДНК, при чему не мењају секвенцу ДНК већ начин на који ћелије користе информације са ДНК. Везујући се за ДНК епигеномска једињења утичу на укључивање/искључивање гена и контролу производње протеина у одређеним ћелијама.

Феном - свеукупност фенотипских особина (уочљивих физичких или физиолошких својстава) ћелије/тквива/органа/организма/врсте.

Транскриптом - скуп молекула РНК произведених у одређеној ћелији/тквиву/организму. Транскриптом организма се активно мења и варира у зависности од многих фактора, укључујући фазу развића и услове животне средине.

Протеом - скуп протеина произведених у организму, систему или биолошком контексту. Протеом није константан; разликује се од ћелије до ћелије и мења се током времена. У одређеној мери, протеом одражава транскриптом.

Метаболом - скуп метаболита малих молекула (<1500 Da) присутних у ћелији/ткиви/организму. Обухвата геномске, транскриптомске и протеомске продукте, као и резултате интеракције геном × средина. Метаболом је веома динамичан јер се мали молекули континуирано апсорбују, синтетишу, разграђују, ступају у интеракцију са другим молекулима, како унутар и између биолошких система, тако и са околином.

Микробиом - скуп свих микроорганизама (бактерије, гљиве, вируси и њихови гени) који природно живе на и у организму.

Фенотипска варијабилност (тј. фенотипске разлике између јединки у популацији) представља скуп различитих фенотипова који развојни систем може произвести систематским варирањем свих својих фенотипских детерминанти - генетичких, еколошких и епигенетичких (Brun-Usan и сар., 2022). У популацији постоји више потенцијалних извора фенотипске варијабилности и сваки од њих је последица различитих основних узрока. Сам извор фенотипске варијабилности одређује да ли дата особина има способност да одговори на деловање селекције (тј. да ли особина има еволуциони потенцијал), као и да ли особина може да одреагује на промене у животној средини (Byers, 2008). Узимајући у обзир различитост фактора одговорних за њен настанак, разликујемо више типова фенотипске варијабилности (Alpert и Simms, 2002):

1. Варијабилност која се објашњава утицајем средине или интеракцијом генотип × средина - односи се на варијабилност познату и као фенотипска пластичност. Фенотипска пластичност представља способност једног генотипа да произведе низ различитих фенотипова у различитим условима животне средине (Pigliucci, 2001). У раним фазама развића организма варијабилност срединских или развојних детерминанти може да утиче на модификовање развојних трајекторија (путања), што резултира и формирањем различитих фенотипова. Дакле, фенотипска пластичност је капацитет једног развојног система да произведе дискретне (полифенизам) или континуиране фенотипске варијације као одговор на варијабилност у окружењу. Однос између срединских варијабли (једне или више варијабли) којима појединачни организам може бити изложен и фенотипских одговора које он производи се приказује нормом реакције (Brun-Usan и сар., 2022). Норма реакције је скуп фенотипова које може произвести појединачни генотип када је изложен различитим условима средине. Фенотипска пластичност може бити адаптивна (пластичност која повећава просечну адаптивну вредност у различитим окружењима, резултат је дирекционе селекције за фенотипску пластичност), штетна (резултат је немогућности одржавања константног фенотипа упркос смањењу адаптивне вредности због варијабилности) или неутрална (услед одсуства селекције за или против варијабилности акумулиране процесом мутације или селекције на другим особинама које су функционално повезане) (Alpert и Simms, 2002).

2. Варијабилност која се објашњава генотипом - односи се на генетички засноване разлике између јединки (генетичка диференцијација, специјализација) и

генетички детерминисане обрасце развића који су константни у различитим срединама (фиксна онтогенетска варијабилност).

3. Варијабилност која се не објашњава ни средином ни генотипом - односи се на варијабилност која је последица развојне нестабилности, тј. степена случајности у развојним исходима. Наиме, развојна нестабилност представља неспособност јединке да произведе специфичан фенотип у датом скупу услова средине (Møller и Swaddle, 1997). Ово одступање формираног фенотипа од циљног фенотипа за одговарајући генотип и окружење представља фенотипску последицу развојног шума који је акумулиран током времена и посредован развојним системом. Развојни шум подразумева случајна варирања у развојним процесима за које се обично сматра да су последица стохастичке природе ћелијских процеса где су кључни молекули (на пример, развојно важне информационе РНК и протеински молекули) присутни у ограниченом (малом) броју у ћелији. Како ће се развојни шум одразити на фенотип зависи од развојних процеса који производе дати фенотип у смислу колико је развојни систем осетљив или робустан (Klingenberg, 2019).

4. Одсуство варијабилности - не постоји увек јасна веза између фактора (варијабилности различитог порекла) који могу да утичу на развојни систем и фенотипске исходе. Некада је упркос изложености различитим факторима фенотипска варијабилност веома мала или потпуно одсуствује. Феномен који описује тенденцију циљног фенотипа да буде константан, иако суочен са унутрашњим (тј. генетичким) и спољашњим (срединским) варијацијама, назива се каналисаност (Klingenberg, 2019). Одсуство фенотипске варијабилности може бити резултат дирекционе селекције против фенотипске варијабилности (адаптивна фиксираност фенотипа), недостатка варијабилности на коју би селекција деловала (генетичко ограничење) или одсуства селекције за фенотипску варијабилност (неутрална фиксираност фенотипа) (Alpert и Simms, 2002).

1.2. Приступуи у квантификовању морфолошке варијабилности

Фенотипска варијабилност се може манифестовати променама у анатомији, морфологији, физиологији, карактеристикама животне историје, понашању и другим својствима организма. Огромна морфолошка разноврсност и сложеност живог света вековима је предмет интересовања научника. Морфологија, научна дисциплина која се бави проучавањем облика, величине, симетрије, регионалне диференцијације и односа делова тела организама првобитно се заснивала на посматрању облика и бележењу метричких запажања. На тај начин научници су настојали да опишу морфолошке и анатомске разлике између организама, разумеју како су разлике настале, утврде везу између облика и функције и пронађу карактеристичне особине (дијагностичке карактере) примењиве у идентификацији врста. Наиме, таксономија и систематика

таксона су се ослањале на визуелни преглед морфологије, квалитативну процену карактера и њихово поређење са референтним примерцима.

Квантитативни приступ у анализи морфолошке варијабилности започиње крајем XIX века са првом применом статистичких метода на морфолошке мере организама. Ипак, развојем модерне статистике и њеном применом у проучавању облика долази до експанзије у области морфометрије средином XX века (Adams и сар., 2013; MacLeod, 2017; Tong и сар., 2021). Морфометрија се бави проучавањем облика организама (у целости или њихових делова), настојећи да прецизно и што потпуније квантифициује информације о величини и облику које су истраживачу често очигледне, али их је тешко адекватно окарактерисати у нумеричком смислу. Стога, морфометријска анализа представља статистичку анализу облика и величине, и њихових коваријација са другим варијаблама (Bookstein, 1991; Dryden и Mardia, 1998). Морфометрија, дакле, обухвата квантитативно мерење и анализу морфолошких особина применом једноваријантних (= униваријантних) и/или вишеваријантних (= мултиваријантних) статистичких метода на произвољне скупове варијабли (тј. морфометријске податке). Морфометријске анализе су првобитно примењиване за анализу морфометријских података попут линеарних дистанци, односа линеарних дистанци и углова (морфометријски приступ назван традиционална морфометрија), али се временом показало да овај приступ има одређене недостатке који ограничавају биолошку интерпретацију варијабилности облика (видети касније). Алтернативни приступ у морфометрији који се у односу на дотадашњи традиционални приступ разликовао у погледу начина квантификавања и анализе морфолошке структуре, свој развој је започео осамдесетих година прошлог века и назван је геометријска морфометрија (Adams и сар., 2013). Због чињенице да користи фундаментално различиту врсту морфометријских података (координате специфичних тачака, контуре, криве, површине) који омогућавају сагледавање геометријских својстава морфолошке структуре и визуелизацију варијабилности облика, геометријска морфометрија се сматра револуцијом у морфометријској анализи (Rohlf и Marcus, 1993; Adams и сар., 2013). За брзи напредак у геометријској морфометрији који је уследио важна је била логична и доследна математичка теорија облика која је, пак, произашла из прецизне дефиниције концепта. Дакле, прецизно и недвосмислено дефинисање облика темељ је математичке теорије облика (Zelditch и сар., 2012). Појам “облик” у геометријској морфометрији се односи на све геометријске карактеристике објекта, осим његове величине, положаја и оријентације (Kendall, 1977; Mitteroecker и Gunz, 2009). Интензивном развоју морфометрије, поред развоја математичке теорије облика, значајно је допринео и развој алата за обраду слике (могућност прикупљања дводимензионалних и тродимензионалних морфометријских података), као и широка употреба брзих персоналних рачунара чиме је анализа и визуелизација високодимензионалних скупова података била могућа.

Неке морфолошке особине се могу успешно квантификовати појединачним мерењем величине њених делова, док се неке друге, сложеније особине не могу окарактерисати само величином, већ су неопходне и информације о пропорцијама и релативном положају делова, тј. о облику (Klingenberg, 2010). Циљ свих морфометријских анализа је да измери морфолошку сличност и различитост, а главна

мера различитости је морфометријска удаљеност (дистанца) која приказује разлику између објеката на основу свих варијабли које су у анализу укључене. У морфометријској заједници воде се дебате о томе који морфометријски приступ употребити, а обично су оне засноване на претпоставци да постоји један исправан приступ у квантификавању облика. Треба имати на уму да сваки од приступа у морфометрији има својих предности и ограничења. Ни један од њих није искључиво једино исправан и одговарајући за све проблеме, већ ће прикладност употребе одређеног начина морфометријске анализе зависити од контекста истраживачког питања које се поставља (Hallgrímsson и сар., 2015). У том смислу, разумевање теоријске основе морфометријских приступа, њихових ограничења и примене је од изузетне важности. У наредним поглављима биће дат преглед типова морфометријских података, основних поставки традиционалне и геометријске морфометрије, те основних метода статистичке анализе.

2. Морфометријски подаци

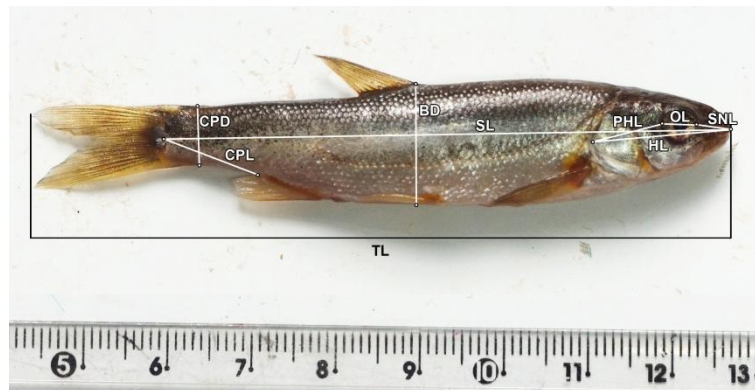
Уопштено говорећи, разликујемо две категорије морфолошких карактеристика:

1. квалитативне (дискретне) особине које карактерише дисконтинуирана варијабилност (стања карактера су јасно раздвојена), а укључују меристичку варијабилност (варијабилност у броју) и квалитативну варијабилност (описна варијабилност);
2. квантитативне (метричке) особине које карактерише континуирана варијабилност (стања карактера се описују њиховим мерама и односима мера).

Постоји више типова података (варијабли) који се користе у морфометријској анализи, а заједничко за све њих је претпоставка истоветности података прикупљених за сваку јединку обухваћену истраживањем. Другим речима, ако мера означена као “стандардна дужина” представља дужину тела рибе измерену од врха носа до њеног последњег пршљена код једне јединке у узорку, онда вредност “стандардна дужина” код осталих јединки мора бити измерена на исти начин; значи, стандардна дужина једне рибе је иста варијабла као и стандардна дужина код друге. Одабир и прикупљање варијабли може зависити од више фактора, почев од постављеног циља истраживања, преко објекта (материјала) који се проучава, доступне опреме за прикупљање података, па све до искуства истраживача.

Дистанце, односи, углови. Дистанце представљају удаљеност између претходно дефинисаних тачака, попут врха носа и последњег пршљена приликом мерења стандардне дужину тела рибе, или на месту одређеном применом неког правила, попут мерења висине тела риба на месту где је она највећа или најмања (Слика 1). Оне могу бити линеарне, праволинијске дистанце, или криволинијске као што су, на пример, крива, обим, контура. По својој природи дистанце су независне од оријентације и

положаја, па сходно раније наведеној дефиницији облика, величина је једина преостала информација коју треба уклонити из скупа дистанци како би се извршила анализа облика, а што се једноставно може постићи употребом односа или индекса добијених из релативних вредности две дистанце (Slice, 2005). Са друге стране, међутим, скуп података сачињен од линеарних дистанци садржи врло мало информација о облику јер се мере могу поклапати, чиме краћа мера може бити обухваћена дужом мером (на пример, дужина главе и дужина тела риба), имати исти смер и полазити из исте тачке због чега међусобно неће бити независне (Слика 1).

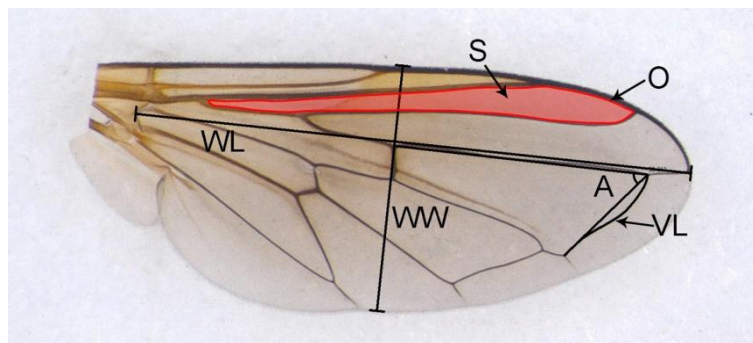


Слика 1. Линеарне дистанце на телу рибе *Telestes dabar* (Leuciscidae) одабране за морфометријску анализу: TL - укупна дужина јединке (од врха носа до краја репног пераја), SL - стандардна дужина јединке (од врха носа до почетка репног пераја), SNL - преорбитална дужина главе (од врха носа до предње ивице ока), OL - пречник ока, PHL - посторбитална дужина главе (од задње ивице ока до најдаље ивице оперкулума), HL - дужина главе (од врха носа до најдаље ивице оперкулума), BD - највећа висина тела (од почетка дорзалног пераја до почетка венталног пераја), CPD - најмања висина тела (измерена на месту где почиње репно пераје), CPL - дужина репног стабла (од краја аналног пераја до последњег кичменог пршљена) (Armbruster, 2012). Извор: модификовано из Francuski и сар. (2019).

Такође, када се одређена дистанца дефинише неким правилом (на пример, висина тела рибе на месту где је она највећа), а не тачно одређеним анатомским тачкама, може се десити да измерене вредности не одговарају датој (кореспондирајућој, хомологној) карактеристици код свих јединки (анатомски гледано место где је висина тела риба највећа може се разликовати између јединки), те се такви резултати тешко могу интерпретирати и упоређивати између врста, па и самих јединки унутар врсте.

У биолошком смислу, **хомологија** представља сличност у структури, физиологији или развићу различитих врста, а која је заснована на њиховом заједничком еволуционом пореклу. У контексту геометријско морфометријске анализе појам хомологије има нешто другачије значење и подразумева да одређене тачке на једној јединци одговарају (кореспондирају) тим тачкама код свих јединки у узорку (тј. представљају “исте” тачке код свих јединки) (Zelditch и сар., 2012). Наиме, пошто се заједничко еволуционо порекло специфичних тачака код анализираних јединки може тешко утврдити, истраживачи овај еволуциони критеријум хомологије (сличност због заједничког порекла) понекад замењују критеријумом подударности (еквивалентности) тачака, а који може бити заснован на анатомији, развићу или функцији (Wärmländer и сар., 2019). Дакле, у морфометријској пракси се избор одговарајућих специфичних тачака заснива на њиховој оперативној хомологији, тј. позиционој подударности (Sneath и Sokal, 1973; Smith, 1990), а претпостављајући да ове тачке имају и биолошку хомологију. Критеријум хомологије приликом одабира специфичних тачака је изузетно важан јер обезбеђује да облик анализиране структуре буде упоредив између јединки/популација.

Додатно ограничење у примени линеарних дистанци је што се у одређеним ситуацијама не могу одредити релативне позиције свих крајњих тачака линеарних дистанци, па се тиме могу изгубити информације о облику анализиране структуре. То се може превазићи додавањем варијабле која ће представљати вредност угла који две дистанце формирају (Слика 2). Попут линеарних мера, и углови су независни од оријентације и положаја, али, такође, и од величине.



Слика 2. Различити типови морфометријских карактера (линеарне дистанце, угао, обим, површина) измерени на крилу врсте *Cheilosia naruska* (Diptera, Syrphidae): WL - дужина крила, WW - ширина крила (измерена дуж линије која спаја anteriornу и posteriornу ивицу крила, а која је постављена на средини и под правим углом у односу на WL меру), VL - дужина вене M_1 , A - угао који затварају крилне вене R_{4+5} и M_1 , O - обим крилне ћелије R_1 , S - површина крилне ћелије R_1 . Извор: модификовано из Milankov и сар. (2010).

Односи или индекси добијени из релативних вредности две линеарне дистанце се понекад сматрају морфолошким својством и узимају за индикаторе разлика у облику структура које се анализирају (Tatsuta и сар., 2018). Додатни опрез потребан је приликом примене мултиваријантне статистичке анализе скупа варијабли које имају различите јединице мерења (линеарне дистанце, односи, углови).

Прикупљање линеарних мера се обично врши директно са објекта истраживања помоћу лењира, мерне траке, нонијуса или неког другог мануелног мерног алата, али развојем компјутерске технологије постали су нам доступни и софтвери помоћу којих се на претходно архивираним дигиталним фотографијама објекта истраживања могу измерити одређене дистанце и углови. У случају када се мерења врше на фотографијама, потребно је објекат истраживања фотографисати уз лењир, милиметарску хартију, или неко друго познато растојање како би се употребом датог стандарда величине одредио фактор размере, а измерене мере биле изражене у дефинисаној јединици дужине. У супротном, измерене вредности ће бити изражене у пикселима. Обележавањем контуре (затворене или отворене) објекта, осим дужине обима могуће је и израчунавање површине унутар дате контуре (Слика 2). Један од програма који се у великој мери примењује за мерење линеарних мера и углова посредством дигиталних фотографија је tpsDig (Rohlf, 2015) (како се користе програми tpsUtil и tpsDig видети у Прилогу 1). Осим tpsDig који је креиран искључиво за прикупљање морфометријских података, и поједини општи програми за обраду слике се неретко користе у ову сврху. Тако, на пример, програм ImageJ (Schneider и сар., 2012) који је намењен да приказује, уређује, анализира, обрађује, чува и штампа дигиталне слике различитих формата (попут tiff, gif, jpg, bmp и других) даје могућност да се измере линеарне дистанце, углови и површине, али и сачувају координате специфичних тачака.

Група програма у оквиру *tps (thin plate spline)* серије намењени су за прикупљање и анализу морфометријских података (Rohlf, 2015). Сви програми ове серије су бесплатно доступни за коришћење у истраживању и настави и могу се преузети на <https://sbmorphometrics.org/>. Серија тренутно обухвата 12 програма, од којих се по својој широкој употреби посебно издвајају tpsDig, tpsRelw, tpsRegr и tpsUtil.

- tpsUtil је изузетно користан програм који има много функција намењених за лако манипулисање датотекама попут креирања датотеке за прикупљање података, промене формата датотеке, брисања или промене редоследа специфичних тачака и јединки, поделе или комбиновања датотека, припреме датотеке неопходне за суперимпозицију клизних специфичних тачака (*semilandmarks*; полутачке), трансформисање контуре/криве у координате специфичних тачака и друге операције. Овим програмом се за одабрани низ дигиталних фотографија анализираног објекта креира почетна *.tps датотека неопходна за следећи корак, односно постављање тачака или прикупљање линеарних мера. Како се користи програм tpsUtil описано је у Прилогу 1.

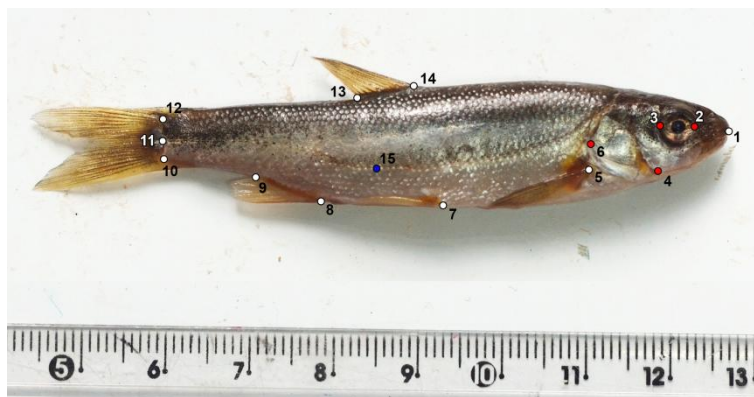
- tpsDig се користи за постављање специфичних тачака и полутачака, крива и контура на дигиталним фотографијама и бележење њихових координата за сваки објекат обухваћен претходно креираном *.tps датотеком. Такође, програм се може користити и за мерење линеарних дистанци, углова, обима и површина. Због чињеница да је лак за употребу и садржи све важне опције за брзо и прецизно прикупљање морфометријских података, готово сви програми специјализовани за геометријско морфометријску анализу облика могу читати податке у формату у којем их овај програм чува (*.tps). Како се користи програм tpsDig описано је у Прилогу 1.
- tpsRelw је програм намењен за суперимпозицију (*superimposition*; процедура којом се уклањања варијабилност условљена разликама у положају, оријентацији и величини; видети касније) конфигурација специфичних тачака и полутачака (скуп координата свих специфичних тачака дефинисаних на структури која се анализира), при чему се израчунавају просечан (консензус, референтни) облик у односу на који се изводи суперимпозиција и центроидне величине конфигурација, а потом изводи анализа главних компоненти варијабилности облика. Програм даје могућност графичког приказивања усаглашене (просечне, консензус) конфигурације и положаја тачака појединачних конфигурација у односу на консензус, као и дистрибуције варијабилности облика у морфолошком простору помоћу графика расејања (*scatterplot*), те промене облика дуж оса главних компоненти. Различите матрице (на пример, консензус конфигурација, конфигурације поравнатих јединки, центроидне величине) могу се сачувати и користити у другим програмима. Како се користи програм tpsRelw описано је у Прилогу 2.
- tpsRegr је програм којим се изводи мултиваријантна регресија варијабилности облика на једну или више независних варијабилности (на пример, величина, температура или неки други еколошки параметар и слично). Такође, као независне варијабилности могу се користити “лажне” (*dummy*) варијабилности којима се кодира припадност јединки различитим групама, тако да се програмом може извести и мултиваријантна анализа варијансе (MANOVA) и мултиваријантна анализа коваријансе (MANCOVA). Ефекат варијабилности независне варијабилности на варијабилност облика је графички приказана помоћу графика расејања и мрежа деформација облика. Како се користи програм tpsRegr описано је у Прилогу 3.
- tpsPls се користи за анализу парцијалних најмањих квадрата (Rohlf и Corti, 2000). Креиран је за проучавање и визуелизацију образаца коваријације између два блока варијабилности прикупљених у истом узорку јединки (на пример, скуп спољашњих варијабилности и варијабилности облика, варијабилности облика две структуре).

Координате специфичних тачака. Координате тачака у правоуглом координатном систему (Декартов координатни систем; систем у коме су координатне осе међусобно нормалне) представљају скуп дистанци, односно удаљености одређених тачака од скупа међусобно нормалних оса система. Када ове тачке представљају анатомске структуре на организму називамо их специфичне тачке (*landmarks*), а врло често су то исте тачке које се користе за дефинисање крајева традиционалних линеарних дистанци или углова (Slice, 2005). Значи, специфичне тачке су дискретне тачке (анатомска места) које се могу недвосмислено препознати као исте тачке (хомологне тачке, кореспондирајуће тачке) код свих јединки укључених у истраживање. За разлику од линеарних дистанци где се свака појединачна мера посматра као особина сама по себи, специфичне тачке се не анализирају и интерпретирају понаособ. Скуп координата свих специфичних тачака дефинисаних на структури која се анализира чини конфигурацију специфичних тачака која садржи информацију о облику структуре и представља један мултидимензионални податак (а не скуп више једнодимензионалних варијабли), те је стога и предмет мултиваријантне статистичке анализе. Наиме, за појединачне специфичне тачке се не очекује да буду биолошки смислене, али конфигурација сачињена од одговарајућег броја пажљиво одабраних и распоређених специфичних тачака ће обезбедити описивање морфологије дате структуре. У успостављеном координатном систему се бележи положај специфичних тачака у виду координата. Применом класичне геометрије и основне тригонометрије на основу добијених координата могу се израчунати све могуће дистанце и углови између постављених специфичних тачака.

Критеријуми за одабир специфичних тачака су (Zelditch и сар., 2012): 1) да позиција сваке специфичне тачке буде јасно дефинисано анатомско место које се у поновљеном постављању тачака може пронаћи више пута и поуздано, у супротном отежана поновљивост и непрецизност могу утицати на појаву грешке мерења; 2) да позиције специфичних тачака буду хомологне (у биолошком или математичком погледу), прецизније речено да специфичне тачке постављене на једној јединци кореспондирају тачкама на свим осталим јединкама у узорку; 3) да распоред специфичних тачака обезбеди адекватну покривеност структуре тачкама, јер не може се детектовати промена облика без података (тј. не може се открити промена на делу структуре која није покривена тачкама); 4) да релативне позиције специфичних тачака буду конзистентне, што се посебно односи на специфичне тачке које могу мењати свој положај једна у односу на другу (на пример, отвор у бази лобање који се налази у близини сугуре који спаја две кости може бити понекад постављен више антериорно, а понекад постериорно у односу на сугуру); 5) да приликом фотографисања јединке буду конзистентно оријентисане испод камере и једна раван буде одабрана за оријентацију објекта, чиме се смањује могућа дисторзија услед пројектовања тродимензионалних објеката у дводимензионални простор и обезбеђује да одабране специфичне тачке леже у истој равни (тј. буду копланарне). Дакле, поштовањем наведених критеријума обезбеђује се да специфичне тачке буду присутне код свих јединки у узорку који се анализира, поновљиве и поуздано постављене, а одабране тако да свеобухватно

представљају морфологију региона или карактеристике од интереса у смислу функције, развића, биомеханике, еволуције и питања која су у истраживању постављена.

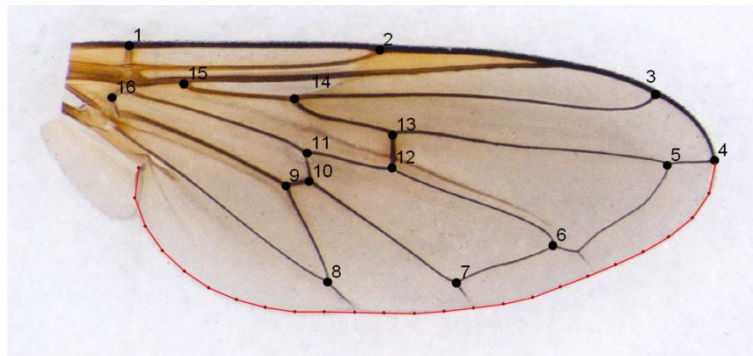
Специфичне тачке се могу класификовати на више начина. Основна разлика се може направити између специфичних тачака (*landmarks*) и полутачака или клизних специфичних тачака (*semilandmarks, sliding landmarks*). У односу на њихову анатомску локацију, биолошки значај и експланаторну улогу, специфичне тачке се даље могу поделити на типове (Bookstein, 1991): Тип 1 су специфичне тачке дефинисане дискретним местом сусрета ткива или структура (на пример, место пресека вена на крилу инсекта); Тип 2 су специфичне тачке чија је хомологија подржана геометријским доказом (на пример, најистуренија тачка на закривљењу); Тип 3 су специфичне тачке чија позиција је дефинисана у односу на друге тачке (на пример, на средини удаљености између две тачке) (Слика 3). Овом дефиницијом за специфичне тачке Типа 3 обухваћене су и клизне специфичне тачке, али њеним каснијим редефинисањем издвојене су још три засебне категорије које обухватају клизне специфичне тачке (Weber и Bookstein, 2011).



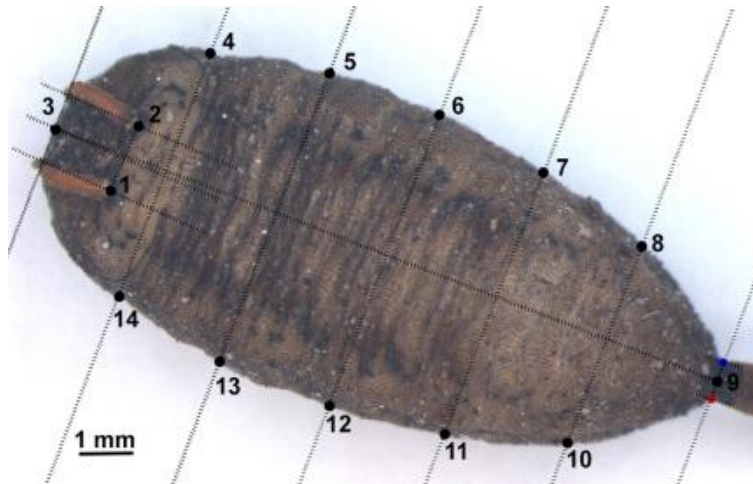
Слика 3. Положај специфичних тачака на телу рибе *Telestes dabar* (Leuciscidae): 1) врх носа; 2) предња ивица ока дуж најдуже осе; 3) задња ивица ока дуж најдуже осе; 4) место највеће закривљености ивице преоперкулума; 5) почетак пекторалног пераја; 6) место највеће закривљености ивице оперкулума; 7) почетак вентралног пераја; 8) почетак аналног пераја; 9) крај аналног пераја; 10) почетак репног пераја вентрално; 11) завршетак кичменог стуба, 12) почетак репног пераја дорзално; 13) крај леђног пераја; 14) почетак леђног пераја; 15) на бочној линији на средини удаљености између специфичних тачака 6 и 11. Специфичне тачке означене белом бојом се могу класификовати као Тип 1, црвеном бојом као Тип 2, а плавом бојом као Тип 3 (Bookstein, 1991). Извор: модификовано из Francuski и сар. (2019).

Оваква подела специфичних тачака је имплицирала и разлике у информативности специфичних тачака, при чему се сматрало да специфичне тачке Типа 1 пружају значајније информације о варијабилности облика јер њихова локација (на пример, на пресеку ткива) подржава хомологију, док, са друге стране, специфичне тачке Типа 3 дају оскудне информације, те су тиме мање поуздане за употребу. Фокусирање на одабир специфичних тачака према типу уноси додатне недоумице и недоследности у вези класификовања тачака. Ипак, све шира употреба методе морфометријске анализе облика засноване на специфичним тачкама указала је да различити типови специфичних тачака дају информације о различитим аспектима облика, те да прецизност мерења више зависи од технике мерења него од типа специфичне тачке. Из тог разлога, одабир конфигурације специфичних тачака треба да буде заснован на њиховој релевантности за тестирање специфичне хипотезе истраживања, а не ком типу припадају (Wärmländer и сар., 2019).

Када се на анализираној структури налазе делови који садрже кључне информације о морфологији на којима одсуствују анатомске локације погодне за постављање специфичних тачака, увођењем додатних тачака може се забележити информација о облику. За разлику од специфичних тачака, ове додатне тачке које се називају полутачке или клизне специфичне тачке (*semilandmarks*, *sliding landmarks*) нису дискретне анатомске тачке и обично се не сматрају хомологним тачкама. Њихово постављање је произвољно, било дефинисањем крајњих тачака на биолошки хомологним местима и алгоритамским распоређивањем одређеног броја додатних тачака између њих (на пример, 10 једнако удаљених клизних специфичних тачака дуж криве; Слика 4), или дефинисањем локација за постављање тачака постављањем референтне мреже преко дигиталне слике објекта (Слика 5). Након постављања додатних тачака, њихове позиције се алгоритамски прилагођавају (видети касније).

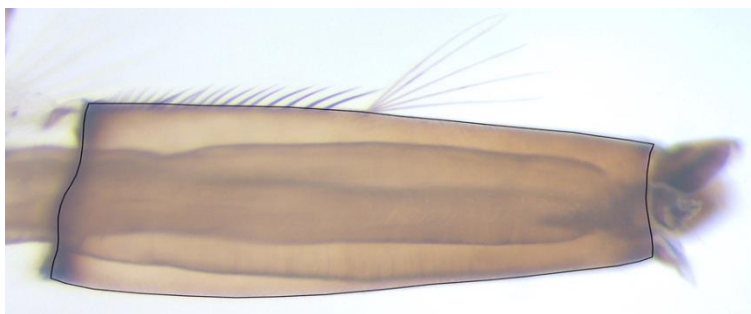


Слика 4. Положај специфичних тачака (тачке 1 - 16) и полутачака постављених на крилу врсте *Cheilosia naruska* (Diptera, Syrphidae). Положај полутачака одређен је помоћу програма tpsDig (Rohlf, 2015) као низ 25 једнако удаљених тачака дуж дефинисане криве (црвена линија). Извор: модификовано из Milankov и сар. (2010).



Слика 5. Положај специфичних тачака (тачке 1 и 2) и полутачака (тачке 3 - 14) постављених на вентралној страни лутке врсте *Eristalis tenax* (Diptera, Syrphidae). Положај полутачака одређен је помоћу референтне мреже постављене преко дигиталне слике програмом MakeFan8 из IMP серије (Sheets, 2014). Мрежу чини серија једнако удаљених и паралелних линија, које су у односу на хоризонталну линију постављене под правим углом. Извор: Ludoški и сар. (непубликовани подаци).

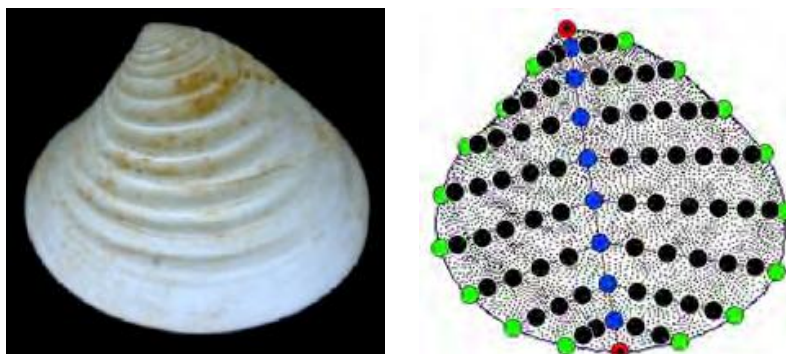
Криве, контуре. Постоје анатомске структуре на којима се не могу обележити добро дефинисане, дискретне и специфичне тачке (или барем не у броју довољном да се забележи варијабилност облика). У том случају, морфометријска анализа облика се може заснивати на подацима добијеним за криве (Слика 4) или контуре (затворена крива) објекта (Слика 6). Ови подаци представљају уређене скупове координата дискретних тачака (углавном клизних специфичних тачака, али могу садржати и специфичне тачке и полутачке) које на први поглед могу личити на податке добијене постављањем специфичних тачака, али су концептуално прилично различити, јер криве и контуре карактерише континуирана структура, а не појединачне тачке. Наиме, координате тачака контуре или криве садрже само информације о „савијању“ криве или контуре, односно степену одступања од праве линије која повезује две експлицитне тачке.



Слика 6. Контура (затворена крива) на респираторној цевчици (сифон) ларве комарца врсте *Ochlerotatus caspius* (Diptera, Culicidae). Извор: модификовано из Francuski и сар. (2021).

Површине. Површине представљају дводимензионалне регионе унутар дефинисаних граница покривене тачкама. Површина објекта се може аутоматски израчунати приликом обележавања контуре или криве на објекту као површина региона ограниченог затвореном линијом, а израчуната вредност је једнодимензиони податак (Слике 2 и 6). У случају када се на објекту обележи крива, софтвер ће спојити почетну и завршну тачку те криве правом линијом, те потом израчунати тако дефинисану површину. Такође, површине се могу узорковати и постављањем скупова тачака. Један од начина постављања скупова тачака илустрован је на Слици 7 примером узорковања површине љуштуре шкољке *Astarte omalii*. Овај начин подразумева да се прво дефинише контура љуштуре, која започиње од специфичне тачке на кљуну љуштуре (црвена тачка; Слика 7). Ова контура садржи низ једнако удаљених полутачака (зелене тачке; Слика 7). Потом је одабрана једна од полутачака на контури (црвена тачка на вентралној ивици љуштуре; Слика 7) да са тачком на кљуну љуштуре дефинише средишњу линију на површини љуштуре (плаве тачке; Слика 7). Од сваке тачке на средишњој линији до тачке на контури по површини љуштуре распоређен је низ међусобно једнако удаљених полутачака (црне тачке; Слика 7) (MacLeod, 2013).

Осим претходно описаног начина, за постављање тачака на површини структуре може се користити мрежа одређених димензија пројектована на површину анализираног објекта (уз неопходни опрез да објекат и мрежа буду исто оријентисани код свих објеката у узорку), где пресек линија на мрежи дефинише локације за постављање тачака (Niewoehner, 2005). Такође, развијају се и нове технике постављања специфичних тачака на тродимензионалне моделе површина и анализе облика површина (Gunz и Mitteroecker, 2013; Boyer и сар., 2015).



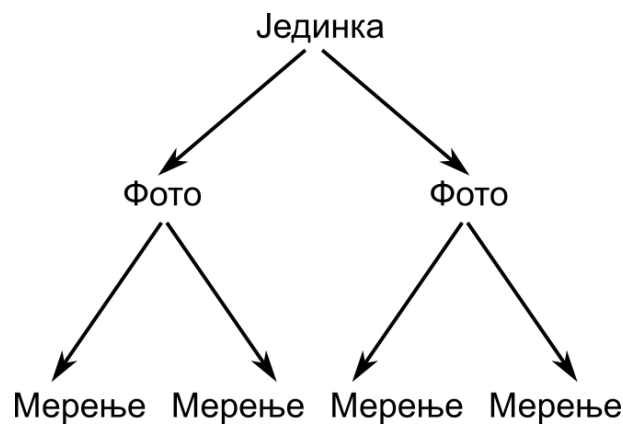
Слика 7. Узорковање морфологије површине љуштуре фосилних представника шкољки врсте *Astarte omalii* (*Bivalvia*, *Astartidae*). Извор: модификовано из MacLeod (2013).

Прикупљање морфометријских података координата специфичних тачака, контура, крива и површина може се извршити помоћу, за ту сврху, дизајниране опреме (на пример, *Microscribe*, <https://gomeasure3d.com/microscribe/>) непосредним бележењем позиције тачака на јединкама или помоћу софтвера (попут програма *tpsDig* и *ImageJ*) који за постављање тачака користе дигитални запис објекта добијен било фото апаратом, површинским и волуметријским скенерима (традиционални оптички скенер, ласерски скенер, компјутерска томографија, магнетна резонанца) или фотограметријом (конструисање тродимензионалних модела на основу дводимензионалних фотографија). Поред одабира позиција специфичних тачака тако да се могу реплицирати од објекта до објекта, а на основу заједничке морфологије, функције или геометрије, важно је да на сваком објекту у анализираном узорку буде постављен исти број специфичних тачака у истом редоследу. Специфичне тачке могу бити дводимензионалне или тродимензионалне и квантификују се у форми $(x,y,[z])$ координата Декартовог система. Употреба програма *tpsDig* у прикупљању морфометријских података мерењем линеарних дистанци, углова и површина, као и постављањем специфичних тачака и полутачака описана је у Прилогу 1.

Неколико проблема се може јавити приликом прикупљања морфометријских података.

Грешка мерења. Мерења (прикупљени морфометријски подаци) су резултат интеракције објекта истраживања, опреме која се користи и истраживача који мерење обавља. Такође, сваки корак током припреме узорака може имати утицаја на исход мерења. Све ове факторе треба узети у обзир као могуће изворе грешке мерења, као и приликом осмишљавања стратегије за минимизирање грешке. Постоје два типа грешке мерења: систематска грешка и случајна грешка (Klingenberg, 2015). Систематска грешка утиче на исти начин на сва мерења која је извршио истраживач са датом опремом; постоји доследно, систематско одступање од стварних вредности. Извори

систематске грешке могу бити разни. Тако, на пример, приликом фотографисања може доћи до дисторзије услед лошег квалитета сочива, или употребе широкоугаоног сочива, паралаксе, неравномерне илуминације различитих делова јединке услед неравномерне осветљености објекта. Приликом позиционирања тачака узроци могу бити деснорукост или леворукост истраживача (посебно у случају примене *Microscribe* опреме), подсвесна пристрасност у позиционирању специфичних тачака на сликама или снимцима (скеновима) леве и десне стране биолошке структуре, неконзистентност у дефинисање позиције тачака (различити истраживачи могу изабрати мало другачије место за постављање исте специфичне тачке). Овај тип грешке може се смањити првенствено одговарајућим подешавањем и калибрацијом опреме. Случајна грешка утиче различито на свако појединачно мерење, а одступања мерења од стварних вредности су у свим правцима (али не нужно у једнаким количинама у свим правцима). Извор случајне грешке може бити присутан у сваком кораку током добијања слика или снимака и позиционирања тачака: позиционирање објекта у односу на камеру (посебно у случају тродимензионалних структура које се анализирају у дводимензионалном простору), варијације у нивоу фокуса или нетачно фокусирање објекта, хетерогена илуминација, недовољна прецизност истраживача приликом постављања тачака и слично. Ефекат овог типа грешке може се ублажити поновљеним мерењима и рачунањем просечних вредности на основу понављања; сви кораци у процесу прикупљања података се могу понављати (Слика 8; Klingenberg, 2015).



Слика 8. Шематски приказ понављања током морфометријске анализе у циљу процене и минимизирања настанка грешке мерења. Свака јединка у узорку фотографише се два (или више) пута уз обавезно поновно постављање јединке испод објектива апарата при сваком фотографисању. За сваку добијену фотографију се два (или више) пута независно изврши прикупљање морфометријских података (постављање специфичних тачака и полутачака, мерење линеарних дистанци и слично). Извор: модификовано из Klingenberg (2015).

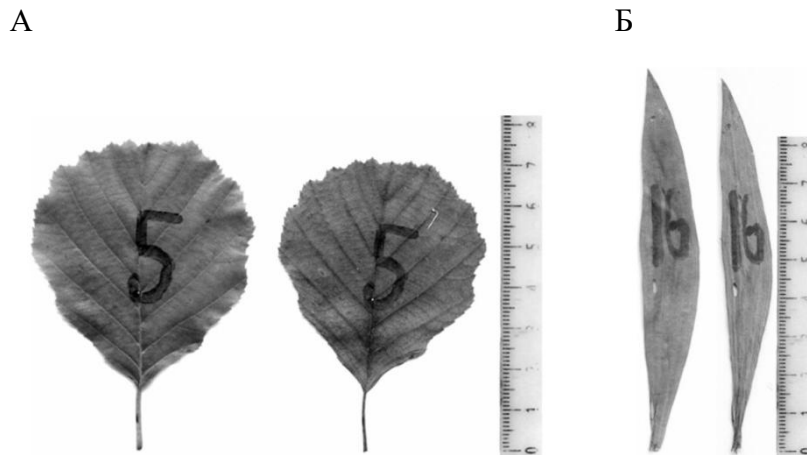
Тако се, у циљу процене величине грешке мерења, препоручује фотографисање истог објекта више пута са његовим репозиционирањем између фотографисања, потом прикупљање података (координате специфичних тачака, контуре, криве) са поновљених фотографија, те утврђивање варијабилности услед грешке у односу на укупну варијабилност у узорку применом статистичких метода. За квантификовање компоненти грешке мерења може се применити Прокрустова анализа варијансе (Прокрустова ANOVA) (Klingenberg, 2015).

Артефакти чувања и складиштења узорака. У неким случајевима, начини чувања и складиштења узорака пре анализе могу утицати на конфигурацију специфичних тачака и тиме произвести варијабилност која није биолошког порекла. Одступање од неутралног или стандардног облика или положаја тела (или делова тела) јединки ће приликом фотографисања и прикупљања морфометријских података неизоставно узроковати додатну варијабилност што се може одразити на анализу облика и значајност статистичких тестова (Arnkvist и Mårtensson, 1998). Тако је, на пример, уочено да код неких врста риба постоји тенденције да им се тело искриви након смрти, да чување узорака риба у замрзивачу такође доводи до искривљења облика тела у односу на стандардни (неутрални) положај (Слика 9), док њихово фиксирање и чување у алкохолу и раствору формалдехида доводи до укочености и/или скупљања тела (Valentin и сар., 2008; Sotola и сар., 2019).



Слика 9. Одступање облика тела риба (искривљеност тела у виду лука) у односу на стандардни (неутрални) положај. Фото: Милица Лукач.

Осим код риба промене у облику и димензијама су утврђене и код других група животиња, као и код биљака. Тако је, на пример, поређењем вредности морфометријских својстава јединки игуана (*Iguana iguana*) измерених непосредно након њихове смрти, а пре депоновања у музејску збирку, са вредностима забележеним за исте јединке два месеца касније уочена промена за поједине карактеристике (Vervust и сар., 2009). Пресовање и сушење биљног материјала узроковало је промене у величини и/или облику појединих биљних структура (Слика 10; Volkova и сар., 2010).



Слика 10. Разлике у величини и/или облику листа пре и након пресовања и сушења биљног материјала. А) Лист врсте *Alnus glutinosa* има мању величину и непромењен облик након препарације. Б) Лист врсте *Plantago lanceolata* има мању величину и издуженији облик након препарације материјала. Извор: модификовано из Volkova и сар. (2010).

Дакле, због свега претходно наведеног, неопходан је додатни опрез приликом прикупљања, нарочито приликом чувања материјала и позиционирања приликом фотографисања, али и анализе морфометријских података. Ипак, неке деформације облика, попут често присутне извијености тела риба у облику лука, могу се кориговати применом операције “*unbend specimens*” у оквиру програма tpsUtil (Rohlf, 2015). Ово је посебно значајно приликом анализе облика фосилних риба код којих се искривљеност положаја тела не може физички исправити.

Недостајуће структуре/некомплетни подаци. Подаци који недостају су још један од могућих проблема са којим се може суочити приликом морфометријске анализе пошто су многе од метода анализе (и традиционалне и геометријске морфометрије) нетолерантне на податке који недостају. Наиме, у компаративним студијама, услед редукције одређене структуре или промена током онтогенетског развоја није могуће код примерака свих врста обухваћених истраживањем позиционирати хомологне специфичне тачке на делу где је дошло до промене. Такође, лоше очувани или

оштећени примерци, нарочито у случају фосилизованих примерака који су често подложни ломљењу, изобличавању и ерозији морају бити изостављени из анализе (Gunz и сар., 2009; Mitteroecker и Gunz, 2009). Проблем недостајућих података се може превазићи смањењем броја специфичних тачака или изостављањем оних мера које се не могу поставити/измерити на свим примерцима, или искључивањем јединке из анализа (Strauss и Atanassov, 2006; Arbour и Brown, 2014). Међутим, када се анализирају фосилни остаци или ретки таксони који су обично доступни у малом броју примерака за анализу, а прикупљање додатног материјала није могуће, изостављање јединки из анализе може додатно компромитовати статистичку анализу података (мала величина узорака, неуједначене величине узорака и слично). Са друге стране, изостављање специфичних тачака и мера може утицати на квантификовање морфолошког диверзитета. Када анализирана структура има билатерну симетрију, положај недостајуће специфичне тачке се може проценити пресликавањем постојеће тачке са једне стране објекта дуж осе симетрије или поравнавање узорка са његовом рефлексијом (Gunz и сар., 2009). Међутим, овај приступ се не може увек применити. Како би се превазишле могуће потешкоће због смањења величине узорка и/или изостављања примерака важних за сагледавање образаца морфолошке еволуције, примењује се неколико метода за процену недостајућих података (Arbour и Brown, 2014; Clavel и сар., 2014). Циљ ових метода није да процени тачну вредност недостајућих података, већ да процене ове вредности на такав начин да се сачувају дескриптивни параметри скупа података, тако да су анализе комплетног скупа података које следе под најмањим могућим утицајем недостајућих података (Clavel и сар., 2014).

3. Линеарна морфометрија

Линеарна, или традиционална како се данас још назива, морфометријска метода примењује униваријантне и мултиваријантне статистичке процедуре на морфометријске податке као што су линеарне дистанце, углови, површине, тежине и односи између варијабли. Морфометријска анализа традиционалних морфометријских података је од великог значаја када нас интересују апсолутне или релативне величине организама (или проучаваних структура), јер оне осим што се користе за представљање величине, могу да одражавају брзину раста и еволуције организама (MacLeod, 2017; Tong и сар., 2021). Међутим, ова метода нема способност да окарактерише целокупан облик организма, а саме варијабле се често третирају као међусобно независне, иако су део веће структуре и самим тим могу да коварирају. Дакле, традиционални морфометријски подаци су мере величине, једнодимензионални подаци који садрже ограничене информације о облику. Пошто не успева да детектује и забележи све геометријске информације о биолошким структурама, линеарна морфометријска анализа не даје потпуну процену варијабилности облика и разлика између анализираних објеката. Непотпуне информације о геометријским односима између анализираних структура могу даље утицати на графичко приказивање и интерпретацију

результата анализе облика. Традиционална морфометрија не прави разлику између величине и облика, те их и не може засебно анализирати.

Примена мултиваријантних метода анализе (попут анализе главних компоненти, каноничке и дискриминантне анализе) на традиционалне морфометријске податке даје могућност да се утврди у ком степену коварирају анализирани морфолошке мере и да се сагледају обрасци варијабилности унутар и између узорака. Међутим, линеарне дистанце су обично високо корелисане са величином (Bookstein и сар., 1985).

Предложено је неколико метода којима се измерене вредности линеарних дистанци могу трансформисати у варијабле кориговане за величину. Тако је један од начина да се измерене вредности линеарних дистанци изразе као процентуални удео дате мере у некој од мера тела (на пример, код риба је обично у односу на стандардну дужину или тоталну дужину рибе).

Такође, апсолутне вредности измерених линеарних дистанци могу да се трансформишу у вредности независне од величине (*size-independent; adjusted*) и помоћу формуле (Elliott и сар., 1995):

$$M_{adj} = M \times (L_s/L_o)^b$$

где је M измерена вредност линеарне дистанце, M_{adj} трансформисана вредност мере M , L_s средња вредност референтне мере дужине у узроку (код риба обично стандардна дужина), L_o вредност референтне мере дужине (стандардна дужина) за дату јединку, коефицијент b нагиб регресионе праве (тј. алометријски коефицијент) утврђен регресијом вредности $\log M$ на $\log L_o$.

Доступни су и други начини за трансформисање традиционалних морфометријских података, али консензус о томе који од предложених начина је најбоље користити није постигнут (Adams и сар., 2004). Пример оригиналних и трансформисаних вредности линеарних дистанци измерених на узорку од 13 јединки риба врсте *Aulopyge huegelii* (Сургинidae) приказан је у Табели 1 (податке уступила Милица Лукач).

Табела 1. Оригиналне измерене и трансформисане вредности одабраних линеарних дистанци 13 јединки риба врсте *Aulopyge huegelii* (Cyprinidae). Трансформисани подаци добијени су израчунавањем процентуалног удела појединачних мера у SL мери и помоћу формуле $M_{adj} = M \times (L_s/L_o)^b$ (Elliott и сар., 1995). За положај измерених линеарних дистанци погледати Сliku 1: SL - стандардна дужина јединке, OL - пречник ока, SNL - преорбитална дужина главе, HL - дужина главе, BD - највећа висина тела, CPD - најмања висина тела. Извор: податке уступила Милица Лукач

Јединка	Оригиналне измерене вредности линеарних дистанци (мм)					
	SL	OL	SNL	HL	BD	CPD
spec01	94,02	5,22	10,63	26,54	23,93	8,28
spec02	111,67	4,90	11,58	28,16	27,40	9,17
spec03	112,44	5,49	12,58	29,89	24,53	9,03
spec04	114,98	5,64	11,29	28,25	27,65	10,09
spec05	96,55	4,91	10,27	25,86	23,16	7,29
spec06	96,84	5,26	9,46	25,94	23,63	8,03
spec07	80,84	4,03	8,15	21,14	20,16	7,24
spec08	124,69	5,89	12,37	31,62	30,15	11,70
spec09	98,44	5,21	10,34	25,64	25,82	8,68
spec10	108,16	4,80	10,59	26,78	24,52	8,99
spec11	104,24	5,27	11,16	27,50	27,24	8,92
spec12	112,44	5,78	12,54	29,68	26,88	9,32
spec13	85,28	4,54	8,30	21,10	20,15	6,92

Табела 1. (наставак)

Јединка	Трансформисане вредности линеарних дистанци (%)				
	OL(% SL)	SNL(%SL)	HL(%SL)	BD(%SL)	CPD(%SL)
spec01	5,55	11,31	28,23	25,45	8,81
spec02	4,39	10,37	25,22	24,54	8,21
spec03	4,88	11,19	26,58	21,82	8,03
spec04	4,91	9,82	24,57	24,05	8,78
spec05	5,09	10,64	26,78	23,99	7,55
spec06	5,43	9,77	26,79	24,40	8,29
spec07	4,99	10,08	26,15	24,94	8,96
spec08	4,72	9,92	25,36	24,18	9,38
spec09	5,29	10,50	26,05	26,23	8,82
spec10	4,44	9,79	24,76	22,67	8,31
spec11	5,06	10,71	26,38	26,13	8,56
spec12	5,14	11,15	26,40	23,91	8,29
spec13	5,32	9,73	24,74	23,63	8,11

Јединка	Трансформисане вредности линеарних дистанци (мм)					
	SL*	OL _{adj}	SNL _{adj}	HL _{adj}	BD _{adj}	CPD _{adj}
spec01	103,12	5,56	11,71	28,88	25,95	9,13
spec02	103,12	4,64	10,65	26,18	25,55	8,43
spec03	103,12	5,17	11,49	27,62	22,74	8,24
spec04	103,12	5,23	10,08	25,57	25,13	8,99
spec05	103,12	5,14	11,00	27,47	24,54	7,82
spec06	103,12	5,49	10,10	27,47	24,97	8,58
spec07	103,12	4,76	10,51	26,41	24,95	9,38
spec08	103,12	5,17	10,14	26,58	25,53	9,56
spec09	103,12	5,38	10,85	26,75	26,89	9,12
spec10	103,12	4,65	10,07	25,64	23,52	8,55
spec11	103,12	5,23	11,03	27,23	26,98	8,82
spec12	103,12	5,45	11,46	27,42	24,92	8,50
spec13	103,12	5,17	10,12	25,10	23,80	8,47

* средња вредност SL мере у узроку

4. Геометријска морфометрија

Имајући у виду сва ограничења употребе традиционалних морфометријских података и линеарне морфометријске анализе у проучавању сложених облика, научници су своју пажњу усмерили на проналажење начина да у анализи облика задрже његов интегритет и геометријска својства што серија линеарних мера није била у могућности. Појава и развој новог приступа у морфометријској анализи названог геометријско морфометријска анализа датира од краја осамдесетих година XX века (Adams и сар., 2004). Геометријска морфометрија је подразумевала нови и другачији приступ како у квантификовању облика морфолошке структуре, тако и у статистичкој анализи података, а све засновано на статистичкој теорији облика која је омогућила комбиновану употребу мултиваријантних статистичких метода и метода за визуелизацију облика (Adams и сар., 2004).

Геометријска морфометрија је мултиваријантна техника која анализира просторну дистрибуцију и варијабилност морфолошких структура на основу конфигурације специфичних тачака (*landmarks*), а која експлицитно раздваја величину и облик помоћу математичких функција као што су трансляција, ротација и скалирање матрица (Bookstein, 1991). Дакле, проучавајући геометрију координата одговарајућих тачака као података који репрезентују морфолошке карактеристике, ова метода у анализу укључује и просторне компоненте варијабилности облика које нису обухваћене традиционалним морфометријским приступом (Adams и сар., 2004; Zelditch и сар., 2012). Наиме, статистичком анализом координата тачака које дефинишу крајеве традиционалних морфометријских мера задржава се комплетна информација о релативном просторном односу података током анализе, што омогућава и визуелизацију групних и индивидуалних разлика. Овакав приступ анализи назива се и морфометрија заснована на специфичним тачкама (*landmark-based morphometrics*). С обзиром да је геометријска морфометрија заснована на положају специфичних тачака и/или полутачака широко прихваћена и популарна метода квантитативне анализе облика, биће у фокусу даљег излагања. Међутим, треба рећи да се осим претходно наведеним приступом облик може описати и на основу затворене или отворене контуре анализираних структуре. Овај концептуално и статистички другачији приступ у геометријској морфометрији, назван и морфометрија заснована на контури (*outline-based morphometrics*), се обично користи у анализи облика структура на којима се не могу (или барем не у довољном броју) поставити специфичне тачке (на пример, љуштура шкољке). Крива (отворена контура) и контура су континуиране структуре дефинисане непрекидним низом тачака чије координате дефинишу степен одступања од праве линије која повезује две експлицитне тачке (тј. савијање криве или контуре). За анализу облика контуре и криве се обично користе Фуријеова анализа (*Fourier analysis*) и еигенанализа облика (*eigenshape analysis*) (за опис и употребу метода видети у Rohlf и Archie, 1984; Rohlf, 1990; MacLeod, 2017; Tatsuta и сар., 2018; Lishchenko и Jones, 2021). У овим анализама се одређена функција (Фуријеов ред, тангентни углови) прилагођава координатама контуре, а израчунати скуп коефицијената који описују

својства дате контуре чини варијабле за наредне анализе. Фуријеова анализа користи Фуријеову декомпозицију за добијање низа тригонометријских функција које описују позиције тачака дуж контуре. Развијено је више метода Фуријеове анализе намењених за отворене и затворене криве/контуре међу којима се елиптична Фуријеова анализа (*Elliptical Fourier analysis*; EFA) сматра најбољом у описивању структуре и посебно ефикасном у редуковању броја варијабли оригиналног скупа података (Lishchenko и Jones, 2021). EFA математички разлаже контуру на синусоидне функције, а еигенанализа облика на ортогоналне функције; објекти су описани као затворене криве x, y координата, а промене у положају x и y су одвојено разложене као периодичне функције (Krieger, 2010). Осим претходно наведеног начина, контура и крива се могу трансформисати у низ тачака (полутачака) чије координате потом чине сет података погодан за анализу облика применом морфометрије засноване на специфичним тачкама.

Упоређујући традиционалну и геометријску морфометрију, могу се издвојити следеће предности геометријске морфометрије: информације о просторном односу између тачака на анализираној структури су садржане у подацима облика, математичко раздвајање облика и величине и њихова независна анализа, могућност непосредне визуелизације облика и просторне локализације варијабилности облика. Такође, стандардизацијом положаја и оријентације конфигурација анализираних тачака из варијабли облика се издваја варијабилност небиолошког порекла.

Због свог потенцијала у превазилажењу недостатака и због ограничења линеарне (традиционалне) морфометријске анализе, геометријска морфометрија је веома брзо своју примену нашла у бројним областима где постоји интерес за квантификавањем морфолошке варијабилности, попут биологије, антропологије, палеонтологије, медицине, инжењерства (на пример, Slice, 2005; Cardillo, 2010; Zelditch и сар., 2012). У биолошким истраживањима, нарочито у контексту систематике, еволуционе и развојне биологије, геометријска морфометрија се користи за процену морфолошких разлика између таксона, сагледавање просторне и временске дистрибуције фенотипског диверзитета, мерење развојне стабилности (флукутирајућа асиметрија), проучавање модуларности и морфолошке интеграције, процену филогенетског сигнала облика и реконструкцију предачког облика, разумевање повезаности облика и функције и друге истраживачке циљеве (на пример, Klingenberg, 2010; Lawing и Polly, 2010; Zelditch и сар., 2012).

4.1. Суперимпозиција

Геометријско морфометријска анализа користи информације садржане у конфигурацији специфичних тачака (конфигурација представља скуп свих специфичних тачака постављених на структури која се анализира). Морфометријски подаци забележени у облику x, y (односно, x, y, z у случају тродимензионалних објеката)

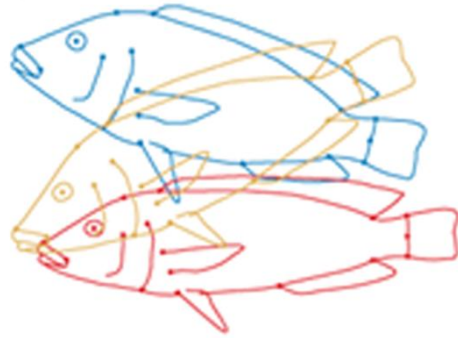
координата у правоуглом координатном систему, а који обично (али не увек) презентују позиције специфичних тачака, међутим, садрже и информације о локацији, оријентацији и величини конфигурације специфичних тачака. С обзиром да, као што је раније већ наведено, облик представља сва геометријска својства објекта изузев његове величине, положаја и оријентације, неопходно је да се ове додатне информације о варијабилности која није у вези са обликом (односно, не утиче на облик, већ само на положај, оријентацију и величину конфигурације специфичних тачака) уклоне. Иако није једина, процедура Прокрустове суперимпозиције (*Procrustes superimposition*) је најчешће примењивана метода којом се информације о варијабилности облика ослобођене варијабилности небиолошког порекла добијају из укупне варијабилности координата специфичних тачака (општи преглед других метода суперимпозиције које се разликују по критеријумима оптималности на којима се заснивају видети у Zelditch и сар., 2012). Укратко, процес Прокрустове суперимпозиције подразумева да се у заједничком референтном систему на све конфигурације специфичних тачака у узорку примене три једноставне операције - скалирање, транслација и ротација (Rohlf и Slice, 1990; Zelditch и сар., 2012; Klingenberg, 2020) (Слика 11).

Први корак Прокрустове суперимпозиције подразумева да се све конфигурације координата специфичних тачака скалирају на стандардну величину. У ову сврху се најчешће употребљава мера величине названа центроидна величина, а која се израчунава као квадратни корен суме квадрата удаљености специфичних тачака од центроиде (= тежишта, центра гравитације) дате конфигурације (Dryden и Mardia, 2016) (Слика 12). Координате центроиде конфигурације специфичних тачака представљају просечну вредност x , односно у координата свих специфичних тачака у конфигурацији. Скалирањем се конфигурације специфичних тачака пропорционално повећавају или смањују како би имале вредност центроидне величине = 1, што се постиже тако што се свака координата свих специфичних тачака подели вредношћу центроидне величине дате конфигурације.

Оригиналне конфигурације
специфичних тачака



Корак 1: скалирање конфигурација
на исту (стандардну) величину



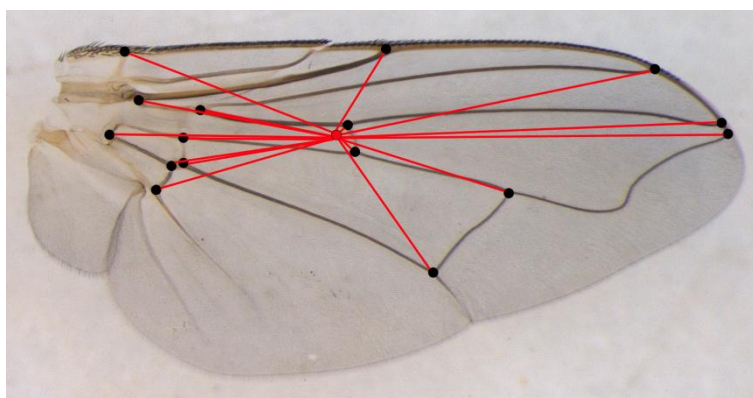
Корак 2: померање конфигурација
на исту позицију



Корак 3: ротирање конфигурација
у оптимални положај (иста
оријентација)



Слика 11. Прокрстова суперимпозиција на примеру три конфигурације специфичних тачака. У кораку 1 оригиналне конфигурације специфичних тачака се рескалирају (пропорционално повећавају/смањују) на стандардну вредност, тј. на вредност центроидне величине 1; у кораку 2 три конфигурације специфичних тачака се померају у координатном систему како би се њихове центроиде нашле у координатном почетку; у кораку 3 конфигурације специфичних тачака се ротирају док не постигну оптималну оријентацију, тј. док се позиције специфичних тачака три конфигурације не поклопе што је боље могуће. Извор: модификовано из Klingenberg (2010).



Слика 12. Центроидна величина представља квадратни корен суме квадрата удаљености специфичних тачака од центроиде дате конфигурације. На слици је приказана конфигурација од 16 специфичних тачака (црне тачке) постављених на крилу врсте *Musca domestica* (Diptera, Muscidae). Положај центроиде (црвена тачка) дате конфигурације специфичних тачака одређен је просечном вредношћу x , односно у координата свих тачака у конфигурацији. Црвене линије приказују растојања која се користе за израчунавање центроидне величине.

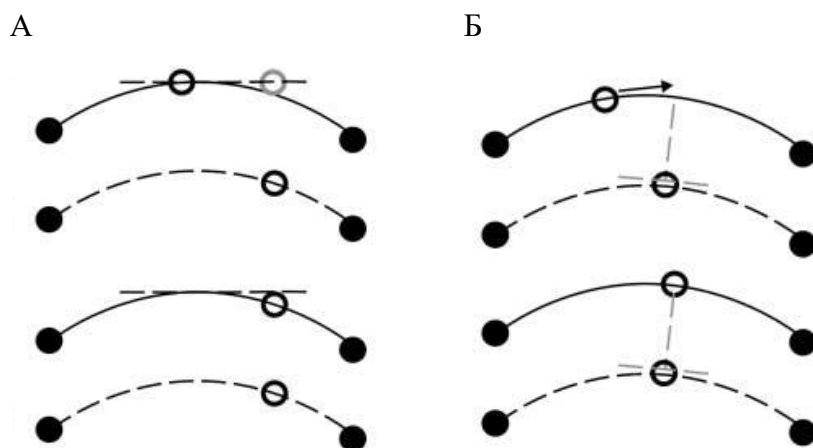
Након скалирања којим се уклањају разлике између објеката условљене величином, конфигурације специфичних тачака се потом померају (транслирају) како би се њиховим постављањем у стандардни положај у координатном систему уклониле варијације у положају (Слика 11). Операцијом транслације конфигурације специфичних тачака се центрирају у координатни почетак (0,0) правоуглог координатног система тако што се координате центроиде дате конфигурације одузму од одговарајућих (x или y) координата сваке специфичне тачке конфигурације, а добијене нове вредности координата специфичних тачака сада рефлектују њихово одступање од центроиде. Центроида конфигурације специфичних тачака се одређује на основу просечног положаја x, y координата.

У последњем кораку суперимпозиције, скалиране и центриране конфигурације специфичних тачака се ротирају око заједничке центроиде све док конфигурације не буду постављене у оптималан положај (Слика 11). Када се пореде две конфигурације специфичних тачака, једна конфигурација се одабира за референтну, а друга се ротира док сума квадрата дистанци између хомологних специфичних тачака између две конфигурације не буде минимална. За више од два облика, овај алгоритам је проширен и назван генерализована Прокрустова анализа (*Generalized Procrustes analysis*; GPA). Све конфигурације специфичних тачака се ротирају док не буду најбоље поравнате са једном, произвољно одабраном конфигурацијом, а потом се израчунава просечан (консензус) облик свих конфигурација који потом представља нови референтни облик у односу на који се ротирају конфигурације до постизања најбољег поравнања. Процес

израчунавања просечног облика и ротирања конфигурација специфичних тачака се понавља више пута, до постизања међусобног поравнања конфигурација које има најмањи могући збир квадрата одступања између хомологних специфичних тачака конфигурација. Операцијом ротације се уклања варијабилност условљена разликама у оријентацији конфигурација специфичних тачака.

Описана процедура суперимпозиције у којој се центроидна величина свих конфигурација специфичних тачака фиксира на вредност 1 назива се парцијална Прокрустова суперимпозиција. Постоји и модификација ове процедуре у погледу скалирања названа потпуна Прокрустова суперимпозиција, где је само циљна конфигурација специфичних тачака скалирана на вредност центроидне величине = 1, док се друга конфигурација скалира тако да одговара циљној што је могуће ближе (за више детаља погледати у Zelditch и сар., 2012 и Klingenberg, 2020). У пракси, разлика између ове две варијанте Прокрустове суперимпозиције обично је веома мала и не утиче на биолошке податке.

Када конфигурације специфичних тачака садрже и координате полутачака (клизне специфичне тачке; *semilandmarks*), стандардна процедура Прокрустове суперимпозиције се мора проширити. Наиме, пошто распоред међусобно једнако удаљених полутачака не подразумева да су њихове позиције хомологне (односно, геометријски или биолошки не кореспондирају) између јединки у узорку (Bookstein, 1997; Gunz и сар., 2005), њима се мора омогућити да се померају (клизе) дуж контуре/криве како би се размак између њих оптимизовао, тј. док не достигну позицију која се најбоље могуће поклапа са позицијама одговарајућих тачака у референтној конфигурацији (другим речима, позиција у којој је сума квадрата одступања између сваке јединке и референтног облика најмања) (Adams и сар., 2004) (Слика 13). Померање ових тачака не одвија се у било ком правцу, већ је ограничено да буде дуж линије тангенте у датој полутачки. Тангента представља линију паралелну са сегментом који повезује низ суседних специфичних тачака или полутачака. Полутачка клизи дуж своје тангенте док се не поравна са вертикалом у одговарајућој полутачки референтног облика. На овај начин се успоставља геометријска подударност клизних специфичних тачака уклањањем ефекта почетног арбитрарног растојања између тачака (Gunz и Mitteroecker, 2013). Најчешће примењивани критеријуми за померање полутачака у процесу суперимпозиције су: 1) критеријум заснован на најмањој енергији савијања који дозвољава да специфична тачка клизи дуж тангенте паралелно са контуром до позиције где је енергија савијања (*banding energy*) неопходна да доведе до промене контуре у односу на референтни облик најмања могућа, и 2) критеријума заснованог на најмањој Прокрустовој дистанци који дозвољава да специфична тачка клизи дуж тангенте до позиције где је дистанца између полутачака појединачних примерака и референтног облика најмања могућа (Perez и сар., 2006) (Слика 13).



Слика 13. Суперимпозиција полутачака применом А) критеријума заснованог на најмањој енергији савијања и Б) критеријума заснованог на најмањој Прокрустовој дистанци. Полутачка (бели кужић) клизи дуж тангенте до постизања оптималног поравнања са кореспондирајућом полутачком референтне конфигурације тачака (референтна конфигурација је приказана црним тачкама повезаним испрекиданом линијом). Извор: модификовано из Perez и сар. (2006).

Након Прокрустове суперимпозиције, конфигурације специфичних тачака које су скалиране, центриране и ротирани садрже само информацију о облику, јер описане операције у процесу суперимпозиције нису утицале на међусобни однос специфичних тачака у конфигурацији, а тиме ни на облик. Такође, координате специфичних тачака у овим конфигурацијама сада имају нове вредности у односу на почетне (односно, пре примене процедуре суперимпозиције) које се називају Прокрустове координате облика и користе се даље у анализи облика. С обзиром да Прокрустове координате описују одступање дате конфигурације специфичних тачака од просечне (консензус) конфигурације називају се и Прокрустови резидуали. Дакле, оно што остаје након овог процеса суперимпозиције је облик сваког објекта у узорку, а оригинална центроидна величина узорака може се издвојити за даљу анализу.

4.2. Простор облика

У геометријској морфометрији, главна мера којом се квантификује разлика између два облика је Прокрустова дистанца. Наиме, Прокрустова дистанца представља квадратни корен суме квадрата удаљености између хомологних специфичних тачака две конфигурације специфичних тачака након Прокрустове суперимпозиције

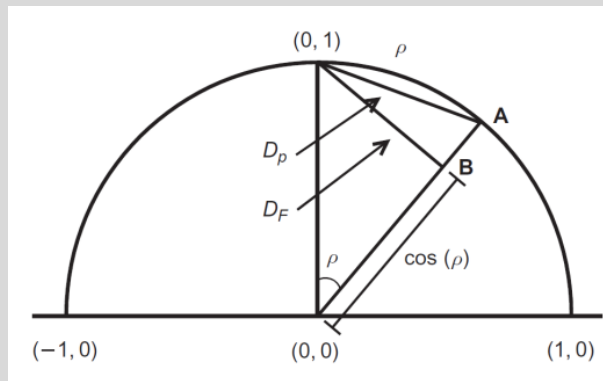
(Mitteroecker и Gunz, 2009). Ова мера може бити парцијална или потпуна Прокрустова дистанца, у зависности од начина скалирања конфигурација који је примењен током процедуре суперимпозиције. Такође, Прокрустова дистанца је еквивалентна растојању између врхова два вектора који садрже све координате специфичних тачака за сваку од две конфигурације специфичних тачака у простору облика. Дакле, варијабилност облика се може окарактерисати у просторима облика. Простор облика је један од неколико видова морфолошког простора који обухвата све могуће облике за одређени број специфичних тачака. У њему је свака појединачна конфигурација специфичних тачака представљена једном тачком, па удаљености између ових тачака указују на сличности између одговарајућих облика, а промене облика су у вези са дистанцама и правцима у простору облика (Klingenberg, 2010). Простор облика је сложени математички простор, који се у геометријској морфометрији назива још и Кендалов (Kendall) простор облика у част шкотског математичара који га је открио и описао његову геометрију. Простор облика није равна површина (тј. линеаран еуклидски простор), већ је он закривљен (нееуклидски простор) и има сложен вишедимензионални облик (за конфигурације од три специфичне тачке то је површина облика сфере, док је за више специфичних тачака он много комплекснији). Димензионалност овог простора зависи од броја специфичних тачака у конфигурацији и димензија података, па тако за дводимензионалне конфигурације специфичних тачака број димензија је $2k-4$, а за тродимензионалне $3k-7$, где k представља број специфичних тачака. Како конвенционални статистички алати претпостављају да је простор линеаран, за даљу статистичку анализу неопходно је да локације објеката у Кендаловом простору облика буду пројектоване у еуклидски простор који представља тангенту постављену на Кендалов простор облика. Тачка постављања тангенте је на референтном облику (обично је то просечан облик целокупног узорка), а пројектовање из једног простора у други се може извршити стереографском или ортогоналном пројекцијом, при чему је због својих геометријских својстава ортогонална пројекција пожељна у већини случајева (за више детаља видети Zelditch и сар., 2012). Овај линеарни простор, назван тангентни простор, је исте димензионалности као и простор облика. Еуклидове дистанце између објеката у тангентном простору су веома сличне дистанцама у Кендаловом простору облика, тако да се облици пројектовани у тангентни простор могу користити за анализу стандардним мултиваријантним статистичким методама.

Морфолошки простори (морфопростори) су математички простори који описују и повезују фенотипове организама. Свака појединачна конфигурација у морфопростору је представљена једном тачком, а геометријски однос између тачака треба да одражава биолошки значајне односе између одговарајућих морфологија (Mitteroecker и Huttegger, 2009).

- Конфигурацију специфичних тачака можемо описати матрицом формата $k \times m$ (тј. матрицом која има k редова и m колона), где k представља број специфичних тачака, а m број димензија (број координата које дефинишу позицију специфичне

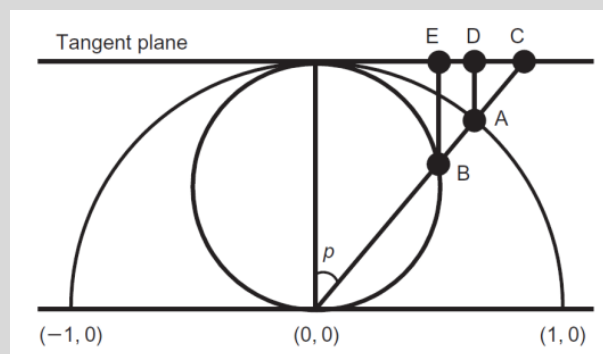
тачке у координатном ситему). Скуп свих $k \times m$ матрица конфигурација специфичних тачака у неком узорку чини **простор конфигурације** (*configuration space*) у коме свака конфигурација може да се прикаже као једна тачка (у овом контексту тачка се односи на појединачни облик) у простору са $k \times m$ димензија.

- Скалирањем конфигурација специфичних тачака на стандардну вредност центроидне величине (центроидна величина = 1) и њиховим центрирањем у координатни почетак (тј. конфигурације специфичних тачака се померају како би њихове центроиде лежале у координатном почетку) елиминишу се разлике у величини и локацији између конфигурација, које сада дефинишу простор који се назива **простор преоблика** (*pre-shape space*; Dryden и Mardia, 1998). Овај закривљени вишедимензионални (хиперсферични) простор има $k \times m - m - 1$ димензија (у односу на број димензија у простору конфигурације губи се једна димензија због скалирања и m димензија због центрирања конфигурација специфичних тачака) и може се замислити као шупља лопта по чијој површини су распоређене тачке које представљају појединачне конфигурације специфичних тачака.
- Након скалирања и translације конфигурација специфичних тачака, једина преостала операција у процесу суперимпозиције је ротирање конфигурација око својих центроида до постизања оптималне оријентације. Оптимална ротација између конфигурација специфичних тачака у простору је она када је Прокрустова дистанца између њих најмања могућа. Оптимално поравнате и скалиране конфигурације специфичних тачака чине **простор облика** чија димензионалност је процедуром ротирања редукована за још једну димензију у односу на простор преоблика. Удаљености између тачака у простору облика одговарају величини разлике између датих облика. Ако се две конфигурације специфичних тачака разликују само по величини, положају и оријентацији, дистанца између њих мора бити једнака нули јер имају исти облик, а тачке које их презентују у простору облика ће се наћи на истом месту. Прокрустова дистанца која се у пракси најчешће користи за квантификовање удаљености између облика се заснива на процедури Прокрустова суперимпозиције (Klingenberg, 2020). Прокрустова дистанца између две поравнате конфигурације скалиране на вредност центроидне величине = 1 назива се парцијална Прокрустова дистанца. Међутим, пошто парцијална Прокрустова дистанца мери удаљеност између два облика па површини сфере, она није најкраћа могућа дистанца између два облика. Ако центроидна величина референтног облика задржи вредност 1 (остаје на површини сфере), а циљног облика промени његовим померањем дуж радијуса ка центру сфере, најмања могућа дистанца између ова два облика ће бити пронађена испод површине простора облика. Центроидна величина циљног облика ће имати вредност $\cos(\rho)$, а дистанца између конфигурација специфичних тачака се назива Прокрустова дистанца (*full Procrustes distance*). Овим скалирањем центроидних вредности конфигурација специфичних тачака у простору облика продукује нову сферу простора облика која се назива **Кендалов простор облика** (*Kendall's shape space*) (Zelditch и сар., 2012; Klingenberg, 2020).



У простору облика поравнатих конфигурација специфичних тачака скалираних на вредност центроидне величине = 1, тетива која спаја референтни (0,1) и циљни облик (тачка А) представља парцијалну Прокрустову дистанцу (D_p). Тачка В је положај циљног облика за вредност центроидне величине $\cos(\rho)$. Линија која спаја референтни (0,1) и циљни облик (тачка В) представља Прокрустову дистанцу (D_F). Извор: модификовано из Zelditch и сар. (2012).

- **Тангентни простор** (*tangent space*) представља линеарни (Еуклидски) простор у који је пројектован простор облика. За тачку у којој тангентни простор додирује простор облика обично се узима просечан (референтни) облика анализираниг узорка.



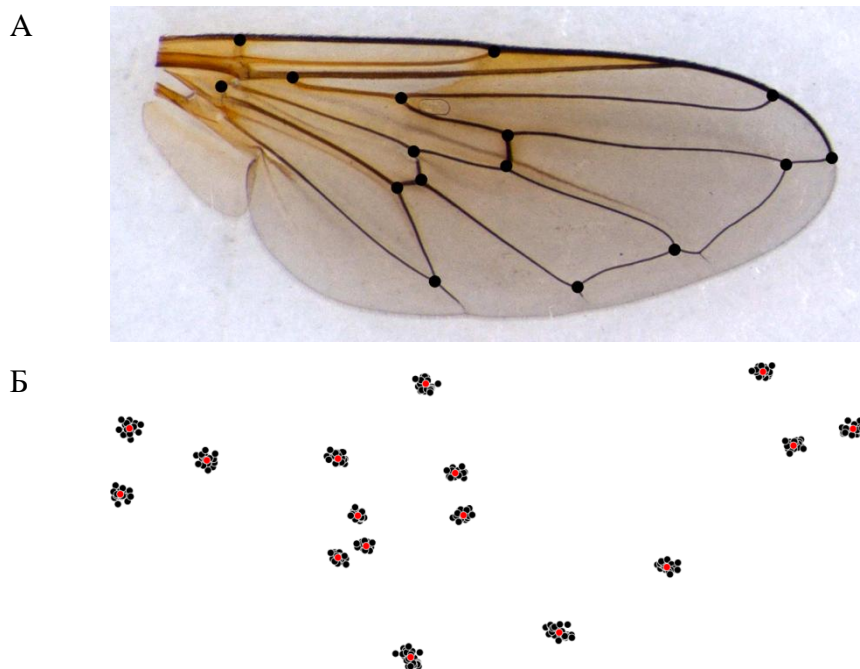
Однос простора облика и тангентног простора. Тачка А представља положај конфигурације облика у простору облика (центроидна величина = 1), а тачка В положај истог облика у Кендаловом простору облика (центроидна величина = $\cos(\rho)$). Тачке D и E су ортогонална пројекција тачака А и В, а тачка С стереографска пројекција тачке А у тангентни простор. Извор: модификовано из Zelditch и сар. (2012).

Као што је раније већ наведено, процедуром Прокрустове суперимпозиције добијене координате облика не садрже никакву информацију о укупној величини (услед скалирања), те се и резултати статистичке анализе ових података могу интерпретирати једино у смислу релативних величина унутар анализираних јединки. Међутим, у многим биолошким истраживањима укупна величина је важан аспект и неодвојив део проучаване морфологије, те је неопходно да информација о величини буде задржана. У том случају, простор облика се може проширити на простор форме тако што се Прокрустове координате облика увећају за вредност центроидне величине дате конфигурације специфичних тачака.

4.3. Визуелизација варијабилности облика

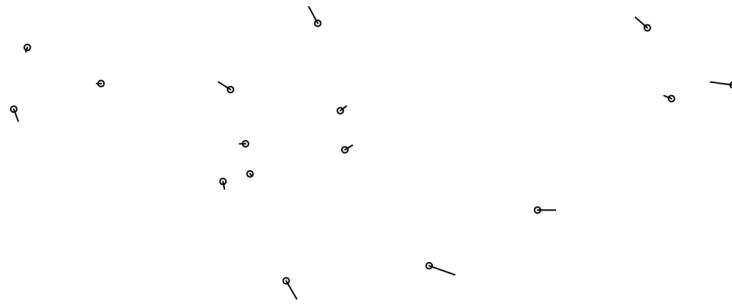
Визуелизација разлика између облика и промене облика су примарни циљ и основна предност геометријске морфометрије. Наиме, након суперимпозиције конфигурација специфичних тачака, израчунавања просечних конфигурација, квантификовања варијабилности и/или утврђивања значајности разлика или корелација, геометријски приступ морфометријској анализи омогућава графичко сагледавање разлика и варијабилности облика. Тиме се резултати статистичких анализа могу интерпретирати и анатомски. Приликом тумачења промена облика мора се имати у виду да се појединачне специфичне тачке не померају независно, већ једна у односу на другу, као и да су промене њихових позиција резултат промена у ткивима или деловима структура где су позициониране. Дакле, графички прикази релативног померања специфичних тачака или деформације дијаграма (на пример, *wireframe* дијаграм) и тродимензионалних површинских модела указују на правац и величину конкретних анатомских промена анализираних структура.

Постоји неколико начина да се визуелизује варијабилност облика (Webster и Sheets, 2010; Zelditch и сар., 2012; Klingenberg, 2013). Најједноставнији начин је путем графика на ком су приказане конфигурације специфичних тачака свих анализираних објеката у узорку након процедуре суперимпозиције (Слика 14). Овакав приказ указује на разлике у позицији кореспондирајућих специфичних тачака између јединки у анализираном узорку. Уколико се овај начин користи за приказивање разлика између група јединки, важно је да се подаци за све групе добију истом методом суперимпозиције, односно у заједничком референтном систему.



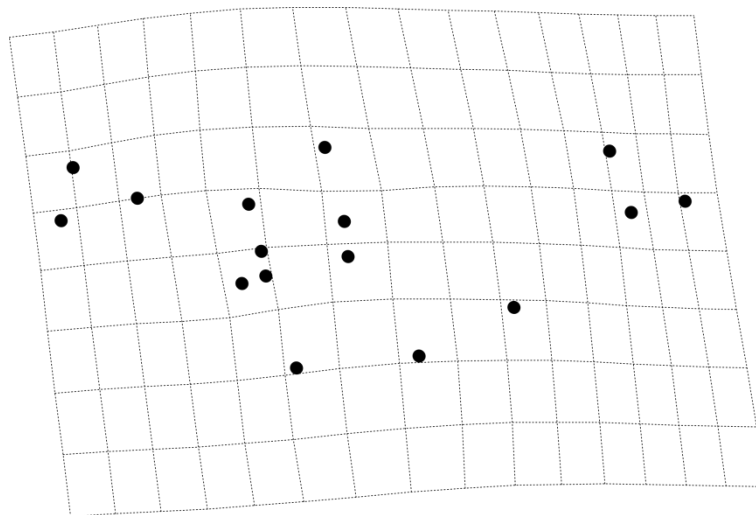
Слика 14. Варијабилност облика крила врсте *Cheilosia naruska* (Diptera, Syrphidae) приказана положајем специфичних тачака појединачних конфигурација након процедуре суперимпозиције (узорак је обухватао 26 јединки). А) Положај 16 специфичних тачака. Б) Положај специфичних тачака у конфигурацији просечног облика приказан је црвеним кружићима, док црни кружићи приказују позиције специфичних тачака појединачних конфигурација.

Разлика у позицији кореспондирајућих специфичних тачака може се приказати и помоћу вектора који повезују положаје специфичних тачака на циљном облику са онима на референтном, при чему циљни и референтни облици могу бити конфигурације специфичних тачака две јединке, или просечне (консензус) конфигурације специфичних тачака две групе јединки које се пореде (Слика 15). Не треба заборавити да поступци који се примењују у анализи облика истражују разлике између облика као целина. У случају визуелизације разлика између облика помоћу вектора, интерпретација резултата у смислу промене позиције појединачних специфичних тачака није оправдана. Наиме, график не садржи информације о томе које специфичне тачке једног облика се налазе на различитој релативној позицији у односу на одговарајуће тачке другог облика (Slice, 2005).



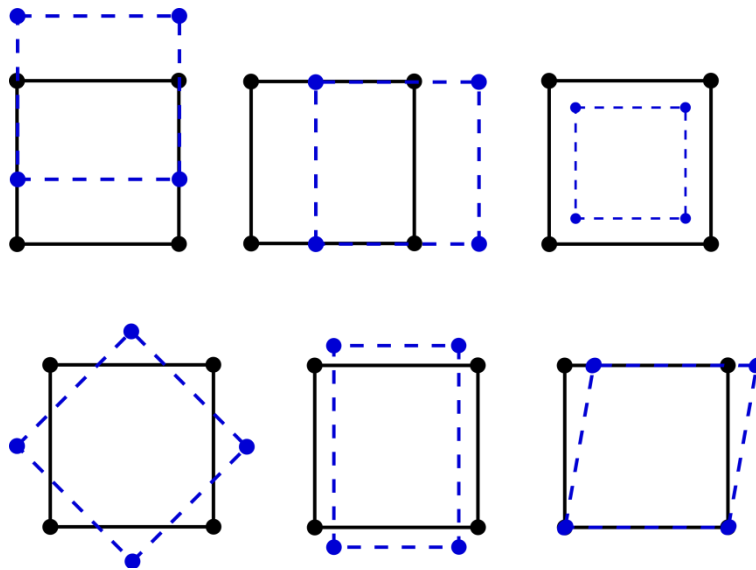
Слика 15. Разлика у просечном облику крила мужјака и женки врсте *Cheilosia naruska* (Diptera, Syrphidae) приказане векторима (промене су увећане 2×). Вектори на слици приказују промену облика крила померањем специфичних тачака (приказаних кружићима) од полазног облика (просечни облик крила мужјака, узорак је обухватао 14 јединки) до циљног облика (просечни облик крила женки, 12 јединки). Положај специфичних тачака приказан је на слици 14А.

Мреже деформација представљају графички приказ који узима у обзир релативне позиције свих специфичних тачака изражавајући разлике у облику одређеног региона анализираних структуре као деформацију истог дела код друге јединке у узроку. Ширење или скупљање мреже показује у којим регионима је код јединке дошло до увећања или смањења у односу на друге јединке (Слика 16).



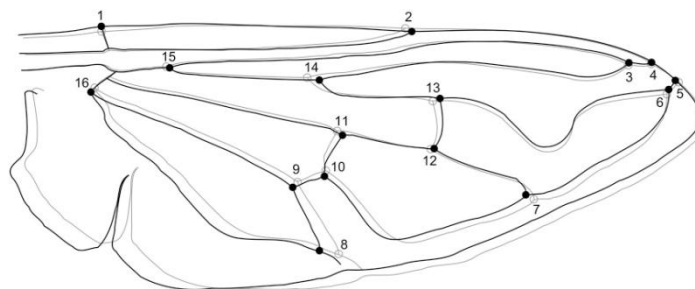
Слика 16. Разлика у просечном облику крила мужјака (полазни облик) и женки (циљни облик) врсте *Cheilosia naruska* (Diptera, Syrphidae) приказана мрежом деформација. Положај специфичних тачака приказан је на слици 14А. Промене су увећане 2×.

Конструисање мрежа деформација врши се применом математичког модела “бесконечно танке металне плоче” (*thin plate spline*; TPS), а који је преузет из физике материјала (Bookstein, 1991). Методом TPS израчунава се функција интерполације која моделује разлику у облику између две конфигурације специфичних тачака минимизирањем енергије савијања потребне да се униформна мрежа (референтна конфигурација специфичних тачака) деформише у конфигурацију специфичних тачака циљног облика (Mitteroecker и Gunz, 2009; Zelditch и сар., 2012). За референтну конфигурацију специфичних тачака се најчешће узима просечан облик добијен након суперимпозиције конфигурација специфичних тачака. Свака координата специфичних тачака анализираних структура може бити више позитивно или више негативно постављена у односу на референтну конфигурацију и ове разлике се приказују као савијање или деформисање мреже. На основу информација садржаних у координатама специфичних тачака, математичким приступом се интерполирају промене облика у регионима између тачака. Дакле, на основу познатих промена у положају специфичних тачака процењују се и реконструирају промене облика у делу између специфичних тачака. Математичка основа за приказивање мреже деформације је метафора савијања идеализоване (тј. бесконачне, униформне и бесконачно танке) металне плоче, према којој се померања специфичних тачака у x, y равни приказује као да се померања дешавају на z осе, односно у трећој димензији (Zelditch и сар., 2012). Количина енергије потребна да се идеализована метална плоча деформише померањем специфичних тачака дуж z осе плоче представља енергију савијања (*bending energy*). Енергија савијања је мера разлике у облику између две конфигурације специфичних тачака, али није метричка мера дистанце и не користи се за статистичку анализу (Mitteroecker и Gunz, 2009). Деформације облика се могу раздвојити на две компоненте: униформну и неуниформну. Униформне деформације обухватају трансформације које не захтевају енергију савијања (транслација, скалирање, ротирање, смицање, компресија/растежање); еквивалентне су нагињању или ротирању плоче тако да паралелне линије остају паралелне (Слика 17). Насупрот униформној деформацији која указује да се иста промена дешава уједначено на целој структури, неуниформне деформације имају локализовани ефекат, односно описују промене облика које имају локацију и просторно су ограничене на регионе у датој структури. Такође, неуниформна деформација се може математички разложити на ортогоналне локализоване компоненте деформације (*partial warps*), а вектори који их мултиплицирају (*partial warp scores*; вектор изражава допринос сваке *partial warp* компоненте у укупној деформацији) се потом могу користити у статистичкој анализи. Сума квадрата *partial warp* скорова једнака је квадрату Прокрустове дистанце између циљног и референтног облика. У већини случајева, биолошке трансформације садрже обе компоненте; комбинација униформних и неуниформних компоненти у потпуности описује неку промену облика. Када су стварне разлике у облику суптилне, мрежа деформација се може екстраполирати произвољним фактором, тако да се увећавањем величине промена може олакшати интерпретација мреже.

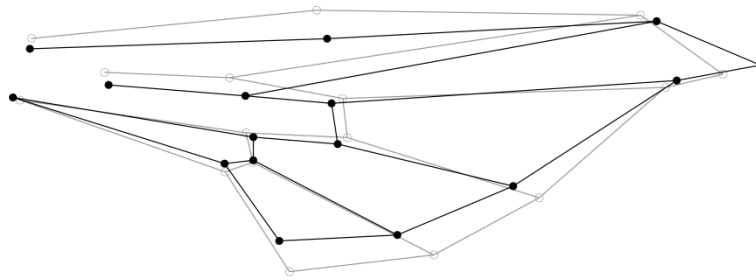


Слика 17. Униформне деформације облика обухватају начине деформација у којима линије које су паралелне пре деформације остају паралелне и након деформације: транслација (померање горе/доле и лево/десно) скалирање (пропорционално увећавање или смањивање), ротирање, компресија или растезање, смицање.

Осим помоћу мреже, промене облика се графички могу приказати и помоћу цртежа контура (*warped outline drawings*; Klingenberg, 2013) (Слика 18) или цртежа са линијама које спајају одабране специфичне тачке (*wireframe*; Klingenberg, 2013) анатомске структуре која се анализира (Слика 19). Ови графици, такође засновани на TPS методи, приказују положај специфичних тачака почетног и циљног облика.



Слика 18. График контуре (*warped outline drawings*) приказује разлику у просечном облику крила мужјака (црна контура) и женки (сива контура) врсте *Eristalis tenax* (Diptera, Syrphidae). Извор: модификовано из Francuski и сар. (2014).



Слика 19. График *wireframe* приказује варијабилност облика крила врсте *Cheilosia naruska* (Diptera, Syrphidae) дуж прве главне компоненте (PC1 оса) у позитивном смеру. Положај специфичних тачака приказан је на слици 14А. Почетни облик (консензус; просечни облик) приказан је линијама сиве боје и белим кружићима, а циљни облик линијама црне боје и црним кружићима.

5. Статистичка анализа морфометријских података

Претходно описане методе које сумирају и визуелизују мултиваријантне податке су углавном истраживачке природе; оне описују податке и обезбеђују нове сетове варијабли (линеарне комбинације оригиналних варијабли) за статистичку анализу података, али не тестирају хипотезе. За тестирање хипотеза се користе неке друге статистичке методе о којима ће бити реч касније.

Статистика је научна дисциплина која истраживачу даје неопходне алате да из прикупљених морфометријских података истражи структуру варијабилности облика, процени значајност разлика у просечним облицима, утврди повезаност утврђене варијабилности са различитим факторима и слично. Варијабла (променљива) представља скуп вредности неког обележја (карактеристике, својства) у узорку (скуп објеката који се анализира). За сваки елемент (јединка, појава) у датом узорку могу бити прикупљене вредности за више од једног својства, тако да скуп података може да садржи више варијабли. На пример, ако за сваку јединку у једном узорку риба измеримо стандардну дужину тела, висину тела, пречник ока и телесну масу, вредности ових мера ће представљати четири варијабле. Методе статистичке анализе које анализирају варијације једне случајне променљиве се називају униваријантним, док се методе које посматрају више повезаних варијабли истовремено називају мултиваријантне.

У традиционалној (линеарној) морфометрији, статистичка анализа се примењује на скуп мерења која обухватају вредности за линеарне дистанце, односе добијене из релативних вредности линеарних мера, углове, масе, запремине, меристичке карактере.

Како ове варијабле често немају заједничку јединицу мерења и упоредиве опсеге варирања, неопходно је да се пре анализе подаци прилагоде и стандардизују. Подаци се могу трансформисати у логаритамске вредности (*log* трансформација) и мултиваријантна анализа се стога обично заснива на корелационој матрици (корелациона матрица је симетрична, елементи испод и изнад дијагонале су вредности коефицијента корелације за све парове варијабли укључене у анализу, на дијагонали је вредност 1 јер је свака варијабла сама са собом у потпуној корелацији).

У геометријској морфометрији, варијабле облика на које се примењује статистичка анализа чине Прокрустове координате облика (добијене у процесу суперимпозиције конфигурација специфичних тачака). Ове варијабле облика имају исту јединицу мерења и међусобно су зависне, због чега су предмет искључиво мултиваријантне статистике (појединачне варијабле облика се не могу анализирати и тумачити засебно, јер све варијабле заједно дефинишу облик), а анализе се заснивају на коваријационој матрици (коваријациона матрица је симетрична, елементи испод и изнад дијагонале су вредности коваријансе за све парове варијабли укључене у анализу, на дијагонали је вредност варијансе). Стандардне статистичке методе се једнако могу применити у анализи облика и на дводимензионалне и тродимензионалне податке. Насупрот варијаблама облика, варијабилност мере величине (центроидна вредност) се анализира униваријантним методама статистичке анализе.

Уопштено говорећи, свака статистичка анализа започиње превођењем биолошке хипотезе у формални статистички (математички) модел. Потом следи израчунавање коефицијената тог модела и тестирање његове статистичке значајности. За разлику од биолошке хипотезе која представља експланаторну изјаву јер садржи и објашњење појаве која се тестира (на пример, облик крила се разликује због варијабилности у величини), математички модели не указују на узрок појаве (на пример, постоји разлика у облику крила).

5.1. Тестирање хипотеза

У току истраживања, истраживач се непрестано сусреће са питањима која се односе на статистичко закључивање: колико су поуздани резултати које је добио и колика је вероватноћа да је разлика између добијених резултата и оних који се очекују на основу постављене хипотезе само последица случајности. Бројни су разлози, од биолошке оправданости до техничких и финансијских, због којих се истраживања не обављају на целокупним популацијама организама, већ на делу популације који се назива узорак. Тако да се и статистичко закључивање о параметрима скупа (тоталне популације) изводи на основу параметара случајно формираног узорка (једног или више узорака).

Поузданост добијених резултата се утврђује одређивањем граница поверења (*confidence limits*; границе поверења представљају доњу и горњу вредност интервала

поверења) за статистичке параметре узорка (аритметичка средина, стандардна девијација, релативна учесталост, параметри регресионе анализе, коефицијент корелације), док се одговор на друго питање добија тестирањем значајности хипотезе. Због чињеница да су многе хипотезе у биолошким истраживањима сложене јер дају више алтернативних теорија о узроку анализиране појаве, при чему те алтернативне теорије не морају међусобно да се искључују и све предвиђају да фактор има ефекат на појаву, подршка статистичких алата у одлучивању између алтернативних хипотеза је неопходна. Хипотеза која се тестира назива се нулта хипотеза (H_0) и представља тврдњу о вредности параметра који се тестира. Насупрот нултој, алтернативна хипотеза (H_1 ; може бити више од једне) обухвата све вредности које параметар може имати, а које нису укључене у нулту хипотезу. Уопштено речено, статистички тест испитује скуп података узорка, а потом се на основу очекиване дистрибуције података, одлучује да ли се може прихватити хипотеза на којој је заснована очекивана дистрибуција или је треба одбацити и прихватити алтернативну.

Током тестирања хипотезе увек постоји ризик да се направи грешка приликом закључивања. Ако је хипотеза тачна, одлука да се прихвати хипотеза биће исправна. Ако је хипотеза тачна, одлука да се хипотеза не прихвати (одбаци) доводи до грешке I типа. Уколико је хипотеза погрешна, одлуком да се она прихвати чини се грешка II типа, а одбацивањем се доноси исправна одлука. Дакле, постоје две врсте исправне одлуке - прихватање тачне хипотезе и одбацивање погрешне, као и две врсте грешке - одбацивање тачне и прихватање погрешне нулте хипотезе. Грешка I типа се обично изражава као вероватноћа и означава са α .

		Нулта хипотеза	
		Прихватање	Одбацивање
Нулта хипотеза	Тачна	Исправна одлука	Грешка I типа
	Погрешна	Грешка II типа	Исправна одлука

Увек постоје неки узорци који су по принципу случајности врло девијантни и могу нас довести у заблуду да верујемо да нулта хипотеза није тачна, те одбацујући је да начинимо грешку I типа. Зато се пре самог статистичког тестирања одлучујемо која величина грешке I типа (α) ће бити дозвољена. Када је грешка I типа изражена у процентима назива се ниво значајности (*significance level*); па тако $\alpha = 0,05$ одговара нивоу значајности од 5% за дати тест. Истраживач може одабрати ниво значајности који жели, али је тај избор за многе тестове ограничен на вредности 0,05 (5%), 0,01 (1%) и 0,001 (0,1%) јер су кумулативне вероватноће одговарајућих дистрибуција углавном доступне у виду табела за ове нивое значајности. Када је нулта хипотеза одбачена за одређени ниво α кажемо да је узорак значајно различит од параметарске или хипотетичке популације (популација из које је узорак узет) са вероватноћом $P \leq \alpha$.

Вредност α већа од 0,05 се не сматра статистички значајном. Ниво значајности од 5% ($P = 0,05$) одговара једној грешци I типа у 20 покушаја/тестова/узорака, а ниво од 1% ($P = 0,01$) једној грешци у 100. Ниво значајности мањи од 1% ($P < 0,01$) се готово увек сматра значајним; нивои између 5% и 1% се могу сматрати значајним по нахођењу истраживача.

Процена значајности. Тест значајности је формална процедура којом се подаци добијени у истраживању пореде са тврдњом тј. хипотезом, чија се истинитост процењује. Циљ примене теста је да се обезбеде докази за одбацавање нулте хипотезе. Ако су докази довољно јаки да одбаце нулту хипотезу, онда се алтернативна хипотеза може аутоматски прихватити. Резултати теста значајности се изражавају у виду вероватноће која указује колико се подаци и тврдња слажу.

Статистичке тестове делимо на параметарске тестове и непараметарске тестове. Класична статистика се ослања на алгебарско извођење формула на основу ограниченог броја добро проучених дистрибуција, посебно нормалне (Гаусове), F , гама, χ^2 , униформне и Поасонове дистрибуције. Параметарски тестови значајности се заснивају на претпоставци да подаци анализирани популације из које је узорак узет прате специфичну дистрибуцију (обично нормалну дистрибуцију), док непараметарски немају овај предуслов. Избор теста ће зависити од постављеног циља истраживања, статистичког модела, структуре података (број варијабли, тип и дистрибуција података) и величине узорка.

Најчешће коришћени параметарски тестови значајности за тестирање нулте хипотезе су тестови који користе t дистрибуцију (Студентов t -тест) и F дистрибуцију (F -тест). Постоји неколико варијанти t -теста: за један узорак, за поређење два узорка и за упарене узорке, а користи се и за тестирање значајности разлика између аритметичких средина ($H_0: \bar{x}_1 = \bar{x}_2$). Вредност t -теста се добија када се разлика између аритметичких средина које се пореде подели са вредношћу стандардне грешке разлике аритметичких средина:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_{\bar{x}_1 - \bar{x}_2}}$$

Израчуната вредност t -теста се потом пореди са граничним вредностима за одговарајући број степени слободе ($df = n_1 + n_2 - 2$, где су n_1 и n_2 величина узорка две групе) и ниво значајности (на пример, $P = 0,05$ или $P = 0,01$), а које се читавају из таблица t дистрибуције. Уколико је израчуната вредност мања од граничне табличне вредности, нулта хипотеза се прихвата као тачна.

Применом F -теста тестира се значајност разлика између аритметичких средина на основу вредности варијанси, односно да ли две популације из којих су узети узорци имају исту варијансу. Дакле, прво се израчунава аритметичка средина сваког узорка, а затим се варијанса између група (одступања аритметичких средина група од укупне аритметичке средине; $MS_{izmedu\ grupu}$) упореди са просечном варијансом унутар сваке

групе (одступања појединачних података од аритметичке средине дате групе; $MS_{unutar\ grupa}$):

$$F = \frac{MS_{između\ grupa}}{MS_{unutar\ grupa}}$$

Ако је дефинисана нулта хипотеза да је вредност аритметичке средине различитих група иста ($H_0: \bar{x}_1 = \bar{x}_2$), онда ће варијанса између група бити иста као и варијанса унутар групе. Израчуната вредност F -теста се упоређује са граничном вредности из табеле F дистрибуције за одговарајући број степени слободе. Број степени слободе за варијансу између група је број група минус 1, а за варијансу унутар група је укупан број података умањен за број група. Ако је гранична вредност већа од израчунате F вредности, нулта хипотеза се прихвата.

Непараметарски тестови се примењују када подаци одступају од нормалне дистрибуције чиме не задовољавају предуслов да се на њих примене параметарски тестови. Треба напоменути да се на податке који одступају од нормалне расподеле може применити трансформација података, попут логаритамске или квадратне трансформације, и тако кориговати ово одступање, те их даље анализирати параметарским тестовима. Такође, непараметарски тестови се користе и за анализу квалитативних података (подаци груписани у дескриптивне категорије) који су по својој природи номинални (групе се не могу поређати по величини) или ординални (рангирана својства), као и када је величина узорка мала. Међутим, иако непараметарски тестови не претпостављају нормалност података, они претпостављају да подаци у различитим групама имају исти облик дистрибуције. Код непараметарских тестова, статистичко закључивање се врши поређењем хипотетичких и из узорка добијених фреквенција, ранга и слично. Међу често примењиваним тестовима (χ^2 -тест, тестови ранга и други) многи од њих се заснивају на χ^2 дистрибуцији. С обзиром на чињеницу да се непараметарски тестови могу применити и на квантитативна својства (нумеричке варијабле) и на квалитативна својства, параметарски тестови имају своје непараметарске алтернативе које се користе у случају да подаци не задовољавају неопходне услове за примену параметарских метода анализе. Ипак, уз све њихове предности, сматра се да је статистичка снага непараметарских тестова углавном мања (када су испуњене претпоставке параметарских тестова), шанса за грешке већа, а приказ и интерпретација резултата су понекад отежани, па их зато треба опрезно користити.

Као што је већ наведено, претходно описане аналитичке математичке методе за тестирање значајности заснивају на одређеним полазним претпоставкама о дистрибуцији података (користе се теоријске дистрибуције). Међутим, када дистрибуција у популацији/узорку није позната, или она значајно одступа од претпоставки аналитичких модела, за процену параметара популације се могу користити модерне методе непараметарске анализе које се заснивају на поновном узорковању (*resampling*). То је група метода која из анализираног узорка, третирајући га као популацију, генерише низ нових узорака (тако симулирајући процес вишеструког узорковање из популације) за које израчунава параметре и формира

њихову дистрибуцију. Дакле, за процену параметара популације користи се дистрибуција добијена емпиријским путем и једина претпоставка коју постављају ове методе је да су подаци добар репрезент популације. Методе засноване на поновном узорку обухватају пермутационе тестове, *bootstrap*, *jackknife*, методе укрштене валидације (*cross-validation*) и Monte Carlo симулације. Сматра се да пермутациони, *bootstrap* и аналитички тестови имају еквивалентну статистичку моћ када подаци задовољавају захтеве аналитичких тестова, али је статистичка моћ тестова заснованих на поновном узорковању већа када нису испуњене претпоставке аналитичких статистичких тестова због чега представљају добру алтернативу аналитичким тестовима (Zelditch и сар., 2012).

Пермутациони тест и *bootstrap* тест се изводе на врло сличан начин и дају сличне резултате, али се разликују у начину на који се формирају нови узорци (који се називају и псеудоузорци). У пермутационом тесту поновно узорковање група је без замене (елемент оригиналног скупа података се у једном псеудоузорку може наћи само једном), што га чини погодним за тестирање хипотеза, али не и за процену интервала поверења (Zelditch и сар., 2012). На пример, ако тестирамо нулту хипотезу да се аритметичке средине величине крила мужјака и женки једне врсте комарца не разликују, вредности измереног својства анализираних јединки две групе чине елементе једног заједничког скупа из кога се насумичним одабиром елемената формирају две групе података (псеудоузорак) који имају исти број елемената као и оригиналне групе, при чему се елементи из заједничког скупа у формираном псеудоузорку не понављају. На овај начин може се формирати велики број псеудоузорака (колико пута ће се поновити формирање псеудоузорака одговара броју пермутација који се у анализи задаје) за које се потом израчунавају аритметичке средине и разлика између ових вредности. Колико пута је разлика између аритметичких средина израчуната за пермутоване податке била већа од разлике добијене за оригиналне податке у односу на укупан број понављања (пермутација) представља вероватноћу да је добијена вредност могла настати случајном поделом једне дистрибуције, или другим речима, разлика између аритметичких средина величине крила мужјака и женки није статистички значајна. Ако је, на пример, од 1000 пермутација разлика између пермутованих сетова била једнака или већа од оне добијене за оригиналне податке у 120 понављања ($P = 120/1000 = 0,120$), не можемо одбацити нулту хипотезу да су средње вредности једнаке за ниво поверења од 5% ($P = 0,05$).

Bootstrap тест се развио из пермутационе и *jackknife* методе. За разлику од пермутационог теста, за *bootstrap* тест се псеудоузорци формирају са заменом тако да се вредност неког елемента оригиналног скупа након што буде узоркована у пермутовани сет података не уклања из оригиналног скупа, па може поново бити узоркована у исти *bootstrap* скуп података. Дакле, псеудоузорци формиран на овај начин су исте величине као и оригинални скупови, али неке вредности из оригиналног скупа података могу бити присутне више пута, а неке ни једном. Даљи поступак се не разликује од оног код пермутационог теста.

5.2. Униваријантне статистичке методе

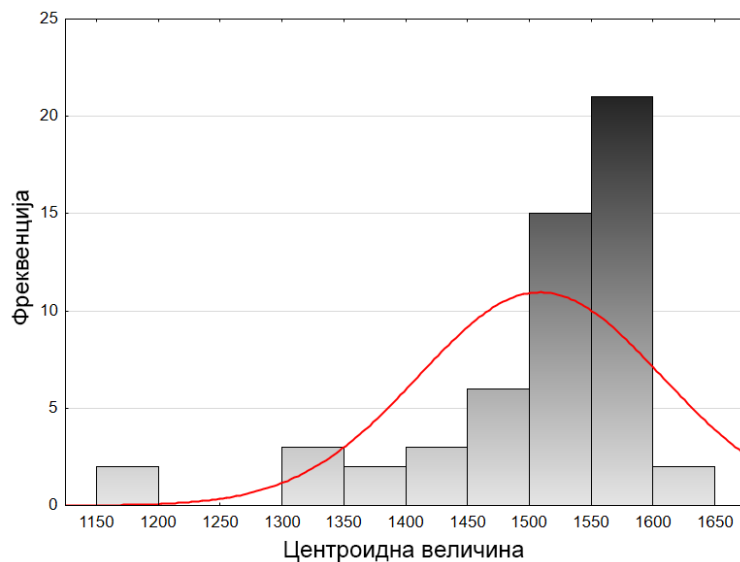
Када су морфометријски подаци за анализиране узорке прикупљени, примарни циљ многих морфометријских истраживања јесте да се применом статистичких процедура на дате податке сагледа структура варијабилности, упореде различити скупови података и процени значајност разлика између њих, сагледа повезаност уочене варијабилности са спољашњих факторима. За ту сврху доступан је читав низ униваријантних и мултиваријантних метода статистичке анализе. Детаљан приказ статистичких метода које се примењују у биолошким истраживањима може се пронаћи у Sokal и Rohlf (1995).

Униваријантне методе се користе за анализу појединачних варијабли (на пример, стандардна дужина тела риба, центроидна величина крила и сл.). Дескриптивна статистика обухвата низ нумеричких и графичких метода којим се вредности анализираниог скупа података групишу и уређују, приказују табеларно и/или графички, те израчунавају основни статистички параметри (мере централне тенденције и мере варијације). Уређивање података подразумева груписање података у статистичке серије на основу вредности анализираниог својства. Тако се, у случају квантитативних карактеристика, вредности скупа података могу систематизовати по величини, а број понављања одређене вредности у скупу података (фреквенција) се наведе уз сваку вредност. Скуп свих фреквенција представља дистрибуцију фреквенција. Такође, дистрибуција фреквенција се може формирати и груписањем вредности датог скупа података у групне разреде. Уређени подаци се осим у виду табела могу приказати и помоћу графикона (хистограм и полигон фреквенције). У примеру у наставку дат је табеларни (Табела 2) и графички (Слика 20) приказ дистрибуције фреквенција вредности центроидне величине крила 54 јединке врсте *Cheilosia naruska*. Центроидна величина израчуната је на основу положаја 16 специфичних тачака (Слика 14А).

Како параметарска статистика претпоставља да су подаци дистрибуирани нормално, пре израчунавања статистичких параметара и тестирања хипотезе неопходно је проверити да ли подаци анализираниог узорка испуњавају овај услов. У ту сврху доступно је више тестова за тестирање нормалности дистрибуције, попут Shapiro-Wilk теста, Kolmogorov-Smirnov теста, Anderson-Darling теста, Lilliefors теста, Jarque-Bera теста и других. Такође, нормалност дистрибуције података се може проверити и на графичком приказу података. Наиме, поређењем хистограма података узорка са нормалном кривом вероватноће, уочава се да ли емпиријска дистрибуција података (хистограм) има облик звона и да ли личи на нормалну дистрибуцију. У примеру из претходног пасуса тест нормалности (Shapiro-Wilk $W = 0,822$; $P < 0,001$) и визуелна процена поклапања криве очекиване нормалне дистрибуције са дистрибуцијом анализираних података (Слика 20) указују да вредности центроидне величине крила анализираниог узорка врсте *C. naruska* одступају од нормалне дистрибуције.

Табела 2. Дистрибуција фреквенција центроидне величине крила врсте *Cheilosia naruska* (Diptera, Syrphidae)

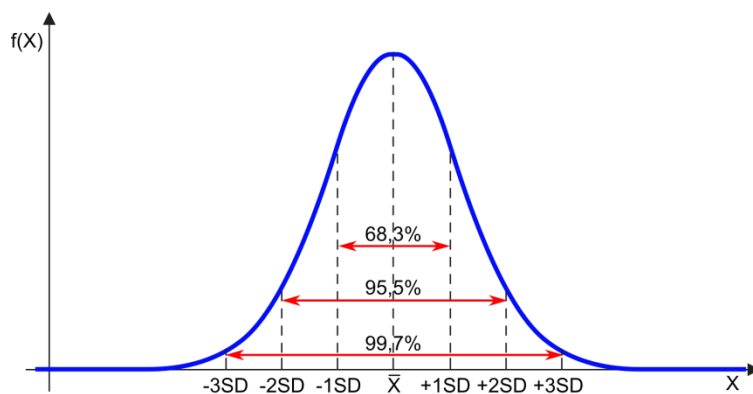
Групни разред	Фреквенција
1150,1 - 1200,0	2
1200,1 - 1250,0	0
1250,1 - 1300,0	0
1300,1 - 1350,0	3
1350,1 - 1400,0	2
1400,1 - 1450,0	3
1450,1 - 1500,0	6
1500,1 - 1550,0	15
1550,1 - 1600,0	21
1600,1 - 1650,0	2
Укупно	54



Слика 20. Хистограм дистрибуције фреквенција центроидне величине крила врсте *Cheilosia naruska* (Diptera, Syrphidae) (подаци приказани у Табели 2). Црвена линија на графику означава очекивану криву нормалне дистрибуције.

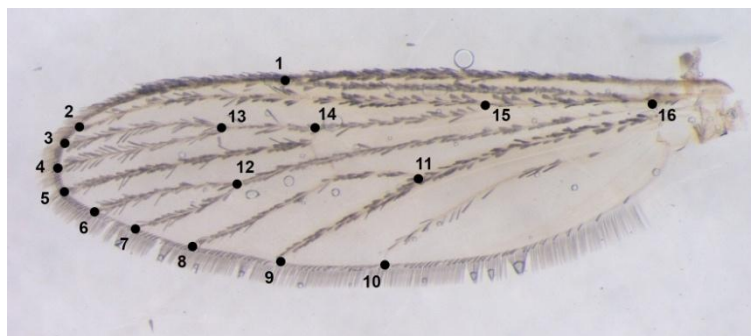
Мера централне тенденције представља статистичку вредност која је заједничка квантитативна одредница свих вредности једне варијабле. Дакле, овим мерама целокупну дистрибуцију можемо изразити само једном вредношћу (бројем). За исказивање средње вредности скупа података најчешће се користи аритметичка средина (просечна вредност која се добија када се сума свих вредности скупа подели са бројем података), која је углавном и основ за тестирање хипотезе и статистичко закључивање приликом поређења више скупова података. Осим аритметичке средине, у морфометријским истраживањима користи се и медијана. Медијана је вредност која се у низу података једне варијабле поређаних по растућој или опадајућој вредности (од најмање до највеће или обрнуто) налази на средини тог низа. Одабир између аритметичке средине и медијане зависиће од природе анализираних података.

Мере варијације су показатељи релативних и апсолутних одступања појединачних вредности скупа података од средње вредности (обично од аритметичке средине) и уз мере централне тенденције дају комплетнију слику о подацима неке варијабле. Другим речима, мере варијације омогућавају да се сагледа облик дистрибуције вредности варијабле; за исту вредност аритметичке средине, облик дистрибуције вредности може се разликовати у зависности од тога каква је варијабилност у скупу података. Најзначајније мере варијације су варијанса (квадратни корен стандардне девијације) и стандардна девијација. На основу стандардне девијације утврђују се интервали поверења тако што се од вредности аритметичке средине одузме и дода једна, две или три вредности стандардне девијације ($\bar{x} \pm 1SD$, $\bar{x} \pm 2SD$, $\bar{x} \pm 3SD$), што је познато као емпиријско правило или правило 68-95-99,7. Ако је дистрибуција података нормална, око 68% података у скупу има вредност аритметичка средина $\pm 1SD$, око 95% вредност аритметичка средина $\pm 2SD$, а око 99,7% вредност аритметичка средина $\pm 3SD$ (Слика 21). Приликом тестирања хипотеза обично се користе стандардни критеријуми интервала поверења од 95% и 99%. Такође, од осталих мера варијације користе се и распон варирања који представља разлику између највеће и најмање вредности скупа података, као и коефицијент варијације. Коефицијент варијације је релативни показатељ који омогућава поређење варијабилности различитих статистичких скупова података (варијабле мерене у различитим јединицама мере). Представља однос вредности стандардне девијације и аритметичке средине датог скупа података, а изражава се у процентима.



Слика 21. Крива нормалне дистрибуције резултата мерења око аритметичке средине (\bar{x}). SD - стандардна девијација.

У Табели 3 приказана је дескриптивна статистика за центроидну величину крила три врсте комараца (*Aedes vexans*, *Anopheles maculipennis* и *Culex pipiens*). Центроидна величина за три анализирани узорка израчуната је на основу положаја 16 специфичних тачака постављених на крилу на местима пресека и завршетка крилних вена (Слика 22). Поред низа основних статистичких параметара који се и најчешће користе поједини статистички програми дају могућност израчунавања и додатних параметара дескриптивне статистике (на пример, модус, хармонијска средина, интервал поверења и други).



Слика 22. Конфигурација специфичних тачака одабрана за геометријско морфометријску анализу варијабилности величине и облика крила врста *Aedes vexans*, *Anopheles maculipennis* и *Culex pipiens* (Diptera, Culicidae).

Табела 3. Дескриптивна статистика за центроидну величину крила врста *Aedes vexans*, *Anopheles maculipennis* и *Culex pipiens* (Diptera, Culicidae)

Параметар	<i>Aedes vexans</i>	<i>Anopheles maculipennis</i>	<i>Culex pipiens</i>
Величина узорка (n)	50	51	50
Најмања вредност (min)	1337,05	1659,94	1643,06
Највећа вредност (max)	1725,40	2318,65	2005,83
Сума (Σ)	76314,43	100457,10	90462,59
Аритметичка средина (\bar{x})	1526,29	1969,75	1809,25
Стандардна грешка аритметичке средине ($S_{\bar{x}}$)	11,47	21,75	10,48
Варијанса (σ^2)	6572,72	24136,38	5488,77
Стандардна девијација (σ ; SD)	81,07	155,36	74,09
Медијана (M_e)	1522,49	1957,89	1806,02
25% перцентил	1478,27	1875,69	1760,87
75% перцентил	1591,37	2092,79	1854,19
Коефицијент асиметрије (<i>Skewness</i>)	-0,08	-0,09	0,37
Коефицијент издужености (<i>Kurtosis</i>)	-0,09	-0,48	0,31
Геометријска средина (GM)	1524,17	1963,68	1807,78
Коефицијент варијације (V)	5,31	7,89	4,09

Израчунавањем статистичких параметара којима се карактеришу скупови података често примећујемо да постоје разлике између добијених вредности. То отвара питање да ли је разлика између две или више аритметичких средина стварна или, пак, резултат случајног варирања у узорку (односно грешке мерења настале услед узимања узорка из популације). Одговор се добија тестирањем хипотезе (видети горе). Када се тестира разлика између аритметичких средина две групе примењује се *t*-тест који има $H_0: \bar{x}_1 = \bar{x}_2$. Вероватноћа да се аритметичке средине два узорка узета из једне популације не разликују (другим речима, да разлика између њих није стварна) може се утврдити применом параметарског приступа, што подразумева поређење израчунате *t* вредности са теоретским прочитаним из таблица *t* дистрибуције за одређени број степени слободе, или непараметарског приступа, односно методама поновног узорковања. Уколико подаци који се анализирају не испуњавају предуслов за примену параметарске статистике (нормална дистрибуција) примењује се Mann-Whitney *U*-тест

(непараметарска алтернатива за t -тест) који тестира хипотезу о једнакости медијана два независна узорка или неки од пермутационих тестова.

Тако се, на пример, на основу параметара дескриптивне статистике приказаних у Табели 3 уочава да се три врсте комараца разликују у величини крила, али не знамо да ли су те разлике статистички значајне. За процену значајности разлике између аритметичких средина центроидне величине крила две врсте, рецимо *Ae. vexans* и *C. pipiens*, користимо t -тест. Пре извођења t -теста, Shapiro-Wilk тестом нормалности је утврђено да у оба узорка дистрибуција вредности не одступа од нормалне. Израчуната t вредност износи $t = 18,219$, а очитана таблична вредност износи $t = 1,9845$ (таблична вредност је очитана за ниво значајности $P = 0,05$ и степени слободe $df = 98$). Поређењем ове две t вредности можемо закључити да је израчуната t вредност већа од табличне и да се нулта хипотеза о једнакости аритметичких средина две врсте може одбацити. Другим речима, разлика између две анализиране врсте комараца у центроидној величини крила је статистички значајна.

За тестирање разлика између три или више независних група примењује се анализа варијансе (*Analysis of variance*; ANOVA). Осим независности узорака, за ову анализу важе и претпоставке о нормалној дистрибуцији података и хомогености варијанси, а тестирање хипотезе се базира на дистрибуцији вероватноће названој F дистрибуција. Уколико су претпоставке нарушене, треба користити непараметарски приступ попут пермутационе ANOVA или Kruskal-Wallis теста за тестирање једнакости медијана. ANOVA је статистичка процедура којом се испитује ефекат једне или више независних варијабли на једну зависну варијаблу. Независна варијабла се назива фактор и он обухвата више група, а у зависности од броја фактора чији утицај на независну варијаблу се процењује, ANOVA може бити једнофакторска или вишефакторска. Током анализе укупна варијабилност (сума квадрата одступања појединачних података од аритметичке средине) се дели на две компоненте: варијабилност унутар узорака (сума квадрата одступања појединачних података групе од аритметичке средине дате групе; сума квадрата унутар групе или SS_{unutar}) и варијабилност између узорака (сума квадрата одступања аритметичких средина група од укупне аритметичке средине; сума квадрата између групе или SS_{izmedu}). Дакле, варијабилност између група је део варијабилности објашњен фактором, а унутар група онај преостали, необјашњени део који се обично назива и резидуал или грешка. Потом се сума квадрата одступања мора поделити са одговарајућим бројем степени слободe како би се израчунала средина квадрата одступања тј. варијанса (MS_{izmedu} и MS_{unutar}). Број степени слободe за SS_{izmedu} добија се тако што се број група умањи за 1, а за SS_{unutar} када се величина узорка умањи за број група. Значајност разлика између група тестира се применом F -теста (видети горе).

ANOVA нам даје уопштен одговор да ли су анализиране групе хомогене, али нам не указује каква је разлика између аритметичких средина појединих група. У случају да тестирањем постављена нулта хипотеза (H_0 : аритметичке средине анализираних група се не разликују) буде прихваћена очекивали бисмо да и разлике између појединих група неће бити значајне. Међутим, може се десити да иако нулта хипотеза не буде одбачена постоје значајне разлике између појединих група, било као

результат случајног варирања или постојања стварних разлика. Управо због тога, потребно је урадити и *post hoc* тест којим се пореде парови група и утврђује значајност разлике између њих. Најчешће у ту сврху корисити се Tukey тест парова.

Разлика у центроидној величини крила између комараца *Ae. vexans*, *An. maculipennis* и *C. pipiens* тестирана је применом једнофакторске ANOVA (у овом примеру фактор чији се ефекат на варијабилност зависне варијабле тј. центроидне величине утврђује је врста). Добијени резултати указују да се три групе (врсте) статистички значајно разликују у величини крила (Табела 4). Такође, применом *post hoc* Tukey теста парова статистички значајна разлика утврђена је и између свих анализираних парова (Табела 5).

Табела 4. Резултати једнофакторске анализе варијансе (ANOVA) центроидне величине крила врста *Aedes vexans*, *Anopheles maculipennis* и *Culex pipiens* (Diptera, Culicidae). *SS* - сума квадрата одступања; *MS* - средина квадрата одступања (варијанса); *df* - степени слободе; *F* - вредност *F*-теста; *P* - *P* вредност (вероватноћа да је тестирана нулта хипотеза тачна; нулта хипотеза се прихвата када је $P \geq 0,05$)

	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P</i>
Између група	5081631,16	2	2540815,58	209,16	< 0,001
Унутар групе	1797831,71	148	12147,51		
Укупно	6879462,87	150			

Табела 5. Резултати *post hoc* Tukey теста парова за центроидну величину крила врста *Aedes vexans*, *Anopheles maculipennis* и *Culex pipiens* (Diptera, Culicidae). Tukey Q вредности су приказане испод дијагонале, а *P* вредности изнад дијагонале

	<i>Aedes vexans</i>	<i>Anopheles maculipennis</i>	<i>Culex pipiens</i>
<i>Aedes vexans</i>	*****	< 0,001	< 0,001
<i>Anopheles maculipennis</i>	28,59	*****	< 0,001
<i>Culex pipiens</i>	18,15	10,35	*****

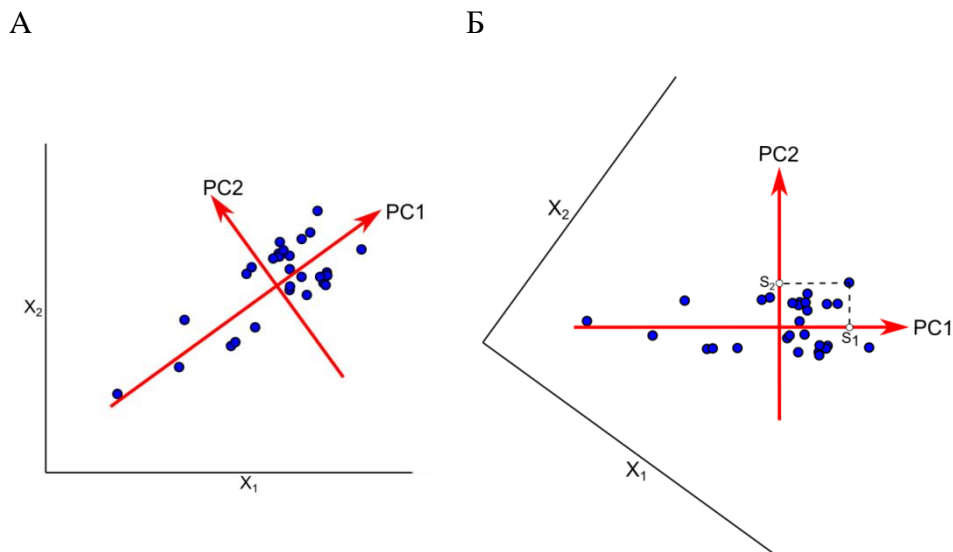
5.3. Мултиваријантне статистичке методе

Опис геометријских својстава објеката обично захтева више од две варијабле, а самим тим и примену мултиваријантне анализе. Методе мултиваријантне статистике омогућавају поређење облика и визуелизацију промена облика. У циљу описивања диверзитета облика и тестирања хипотеза међу најчешће коришћеним су анализа главних компоненти, каноничка анализа (дискриминантна анализа), мултиваријантна анализа варијансе, кластер анализа, регресиона анализа и анализа парцијалних најмањих квадрата.

Услед високе димензионалности конфигурација специфичних тачака простор облика је тешко визуелизовати и тумачити. Варијабилност у простору облика може се сажети и поједноставити применом метода мултиваријантне статистике које описују диверзитет облика помоћу новог (мањег) скупа варијабли изведеног из великог броја оригиналних варијабли на такав начин да се укупна варијабилност и коваријабилност у скупу података не изгуби. Дакле, ове методе (називају се и методе ординације; *ordination methods*) сумирају варијабилност у простору облика, омогућавају графичко приказивање и визуелно уочавање образаца, али не и тестирање хипотезе. Из ове групе метода обично се користе анализа главних компоненти и каноничка анализа.

Анализа главних компоненти (*Principal component analysis*; PCA) је метода којом се велики скуп оригиналних међусобно зависних (корелираних) варијабли редукује на мањи број димензија које и даље обухватају велики део укупне варијабилности података. Нови скуп варијабли ће чинити линеарне комбинације оригиналних варијабли које имају максималну варијансу и међусобно су независне. Нове варијабле називају се главне компоненте. Наиме, анализа главних компоненти подразумева разлагање матрице коваријације узорка и израчунавање нових варијабли чије варијансе када се сумирају имају исту укупну вредност као и вредност варијансе и коваријансе оригиналних варијабли, док је коваријанса свих главних компоненти једнака нули. Прва главна компонента садржаће највећи могући део укупне варијансе, а свака наредна обухватиће највећи могући део варијансе која није описана претходном компонентом (Слика 23А). Број генерисаних главних компоненти биће једнак броју оригиналних варијабли облика, при чему обично највећи део укупне варијабилности у узорку може бити описан са само неколико компоненти. Главне компоненте, потом, се могу користити као основа за нови координатни систем чинећи осе тог система, док се позиција јединки у њему утврђује ортогоналном пројекцијом јединки на главне компоненте (Слика 23Б). Пројекцијом добијене дистанце (удаљености између места пројекције и места пресека две главне компоненте) називају се скорови главних компоненти и представљају удаљености јединки од средње вредности узорка (јер се главне компоненте секу у месту где се налази средња вредност узорка) у правцима датих компоненти (Слика 23Б). Дакле, анализом главних компоненти се генерисањем нових варијабли, оригинални скуп података ротира, при чему положај тачака (јединки) у међусобном односу остаје непромењен у односу на онај у оригиналном координатном

систему, па оригиналне податке можемо видети из другачије перспективе (Слика 23А, Б).



Слика 23. Анализа главних компоненти. А) График расејања приказује дистрибуцију јединки (плави кружићи) у морфолошком простору дефинисаном варијаблима X_1 и X_2 . Правци који пролазе кроз скуп података (црвене стрелице), а који описују највећи део укупне варијабилности називају се главне компоненте (PC1 и PC2). Б) Скуп података се ротира, а главне компоненте (PC1 и PC2) постају осе новог координатног система. Тачке S_1 и S_2 су места ортогоналне пројекције јединке на PC1 и PC2 осе, а њихове удаљености од места пресека главних компоненти представљају скорове главних компоненти.

Резултати анализе главних компоненти (као илустрација узет је пример анализе варијабилности облика крила комараца *Ae. vexans*, *An. maculipennis* и *C. pipiens* методом геометријске морфометрије; конфигурација специфичних тачака приказана је на Слици 22) садрже листу генерисаних главних компоненти и вредност варијансе сваке компоненте (еигенвредности), проценат укупне варијансе објашњен сваком компонентом (Табела 6) и скорове главних компоненти за сваку јединку за сваку компоненту (Табела 7).

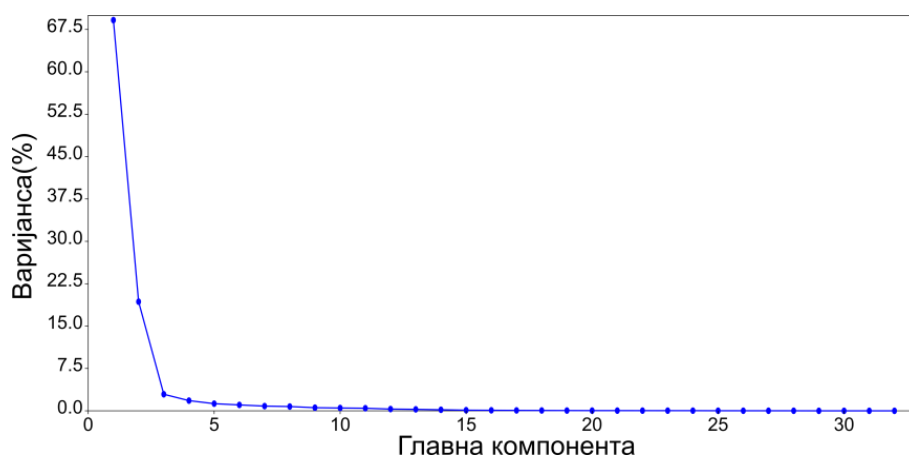
Табела 6. Резултати анализе главних компоненти варијабли облика (Прокрустове координате) крила врста *Aedes vexans*, *Anopheles maculipennis* и *Culex pipiens* (Diptera, Culicidae)

Главна компонента (PC)	Варијанса	% варијансе	Кумулативни % варијансе
1	0,00302806	69,151	69,151
2	0,00084582	19,316	88,467
3	0,00012833	2,931	91,398
4	0,00007907	1,806	93,204
5	0,00005546	1,267	94,470
6	0,00004568	1,043	95,513
7	0,00003661	0,836	96,349
8	0,00003319	0,758	97,107
9	0,00002427	0,554	97,662
10	0,00002188	0,500	98,161
11	0,00002007	0,458	98,620
12	0,00001359	0,310	98,930
13	0,00001150	0,263	99,193
14	0,00000810	0,185	99,378
15	0,00000493	0,113	99,490
16	0,00000411	0,094	99,584
17	0,00000329	0,075	99,659
18	0,00000277	0,063	99,723
19	0,00000217	0,050	99,772
20	0,00000175	0,040	99,812
21	0,00000168	0,038	99,850
22	0,00000150	0,034	99,885
23	0,00000135	0,031	99,916
24	0,00000131	0,030	99,945
25	0,00000078	0,018	99,963
26	0,00000065	0,015	99,978
27	0,00000052	0,012	99,990
28	0,00000045	0,010	100

Табела 7. Индивидуални скорови главних компоненти добијени анализом главних компоненти варијабли облика (Прокрустове координате) крила врста *Aedes vexans*, *Anopheles maculipennis* и *Culex pipiens* (Diptera, Culicidae). Приказане су вредности скорова првих пет (од 28 генерисаних) главних компоненти (PC1 - PC5) за десет (од 150) анализираних јединки

Јединка	PC1	PC2	PC3	PC4	PC5
1	0,020862	-0,031985	-0,005316	-0,001826	0,014599
2	-0,005735	-0,037498	0,006266	-0,000682	-0,000365
3	-0,013912	-0,046197	0,012588	-0,014899	0,003445
4	0,019691	-0,050660	-0,020384	0,008176	0,002595
5	0,015400	-0,038966	-0,000644	-0,007337	0,003442
6	0,026952	-0,038690	-0,006965	-0,011334	-0,005991
7	0,021941	-0,034306	-0,015039	-0,003131	-0,015348
8	0,009572	-0,029952	0,004607	0,005468	0,001462
9	0,008464	-0,047780	-0,004889	-0,001612	0,008197
10	0,012395	-0,027414	0,010898	-0,003182	-0,000628

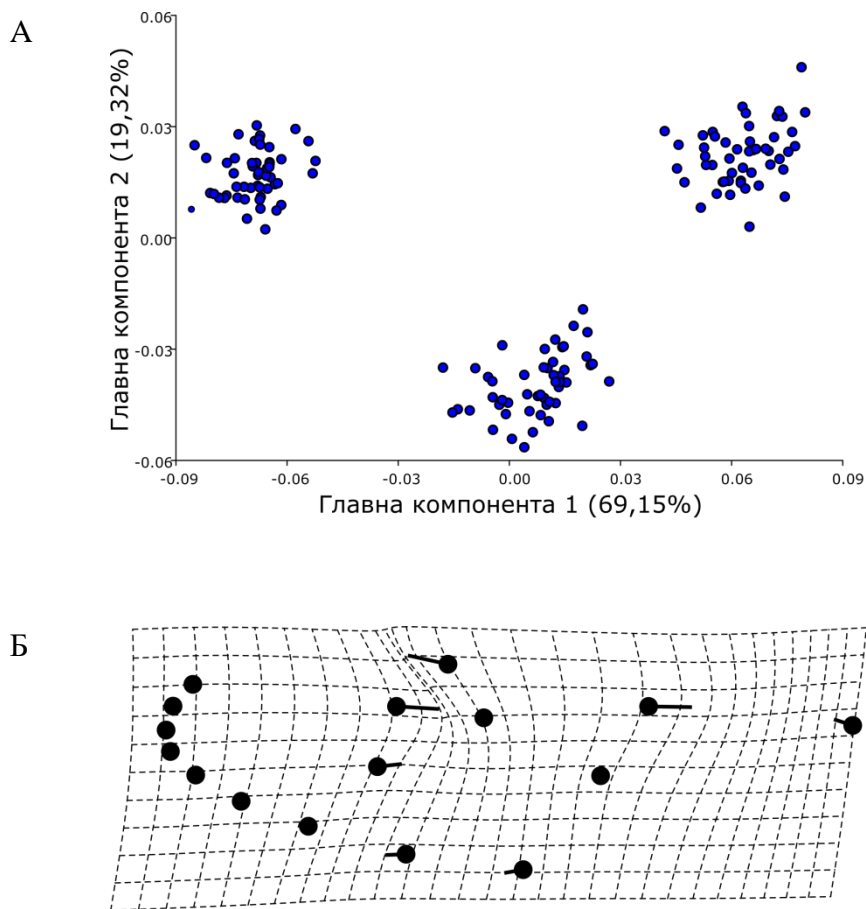
С обзиром да варијабилност коју објашњавају главне компоненте опада са сваком наредном компонентом и да је веома често највећи део укупне варијабилности података обухваћен са првих неколико компоненти, сматра се да приликом интерпретације добијених резултата компоненте које носе мали део варијабилности (једно од често примењиваних правила каже мање од 5% укупне варијабилности) није неопходно разматрати. Број главних компоненти које ћемо интерпретирати можемо одредити и помоћу графичког приказа на коме су вредности варијансе или проценти укупне варијансе приказани у односу на редни број главне компоненте (*scree* график) (Слика 24). На графику се препознаје место где се крива која спаја ове вредности нагло мења (“ломи”) јер две узастопне компоненте имају различите варијансе (у PC3 на Слици 24), па се интерпретирају само оне главне компоненте лево од те тачке. Дакле, осим што представља брз и погодан начин да се проуче подаци, анализа главних компоненти пружа могућност да се редукује димензионалност података. Другим речима, главне компоненте које носе највећи део варијабилности се могу користити као улазни сет података за друге мултиваријантне анализе (на пример, узме се сет компоненти које носе укупно 95% или 99% варијансе).



Слика 24. Графички приказ вредности варијансе описане главним компонентама (вредности варијансе су приказане у Табели 6). Место где се крива која повезује вредности варијансе (плава линија) нагло мења (“ломи”) је у главној компоненти 3 (PC3).

Скорови главних компоненти могу се графички приказати у виду дводимензионалног или тродимензионалног графика расејања (*scatterplot*) који приказујући како је дистрибуирана варијабилност у морфолошком простору дефинисаном главним компонентама могу указати и да ли у анализираном узорку постоји груписање јединки са сличном морфологијом, а да претходно није дефинисана припадност јединки одређеним групама (Слика 25А). Такође, када се оригиналне варијабле облика помноже са коефицијентима главних компоненти и сумирају, може се визуелизовати и варијабилност облика дуж неке од главних компоненти (Слика 25Б).

У многим истраживањима анализа главних компоненти када се примени на геометријске податке назива се *relative warp* анализа и уколико варијабилност није пондерисана са енергијом савијања ове две анализе су у потпуности еквивалентне.

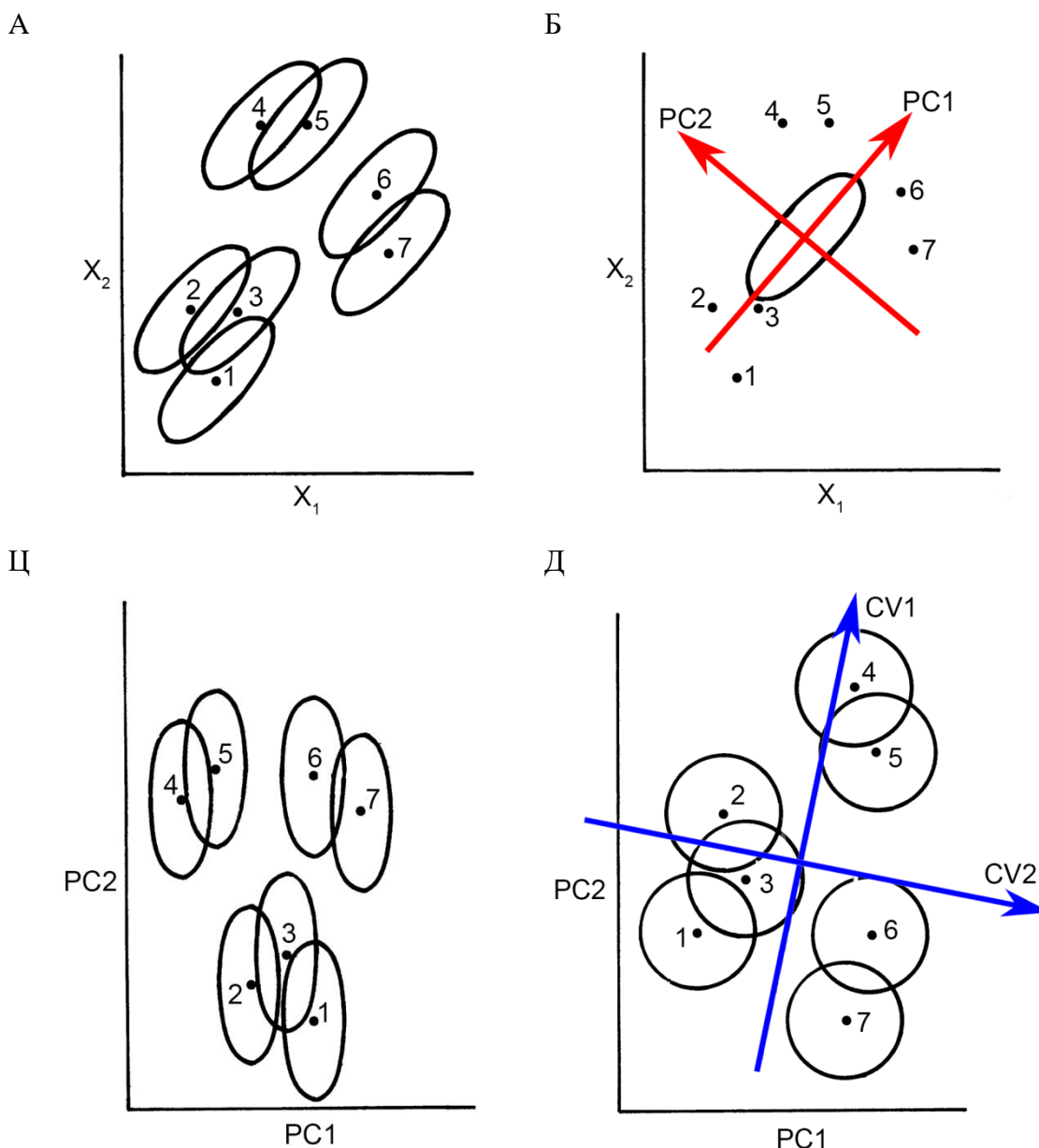


Слика 25. Графички резултати анализе главних компоненти варијабли облика (Прокрустове координате) крила врста *Aedes vexans*, *Anopheles maculipennis* и *Culex pipiens* (Diptera, Culicidae). А) График расејања приказује дистрибуцију јединки у морфопростору дефинисаном са прве две главне компоненте. Б) Мрежа деформација облика крила дуж прве главне компоненте. Положај специфичних тачака у конфигурацији просечног облика (консензус) приказан је црним кружићима, а векторима (црне линије) релативна промена њиховог положаја са повећањем скора главне компоненте.

Каноничка анализа (*Canonical variate analysis*; CVA) поједностављује описивање разлика између група јединки и утврђује математичке функције којима се одређује припадност јединки дефинисаним групама, али не тестира разлике у просечним облицима. Дакле, ова метода се у морфометријским истраживањима примењује у циљу дискриминације (разликовања) група јединки и идентификовања које варијабле највише доприносе разликама између група, описивања морфолошких разлика и мерења укупне разлике између група, те утврђивања у коју од претходно дефинисаних група би се могле разврстати неklasификоване („непознате”) јединке. Пошто CVA подразумева постојање група и да су разлике између њих статистички значајне, стога је пре саме примене ове анализе неопходно дефинисати групе јединки. Јединке се сврставају у исту групу ако деле исто стање неког дисконтинуираног својства (на пример, истог су пола, припадају истој врсти, сакупљене су на истом локалитету и слично) и по датом својству могу припадати само једној од група.

Примена мултиваријантних техника класификације на морфометријске податке је веома честа у бројним биолошким дисциплинама (попут систематике, екологије, палеонтологије, антропологије, форензике и других). Није реткост да се каноничка анализа назива и дискриминантна анализа (*Discriminant analysis, Linear discriminant analysis*). Иако се ови називи данас углавном користе као синоними, треба нагласити да је дискриминантна анализа првобитно била развијена за анализу две групе, а потом је проширена за анализу три или више група и названа каноничка анализа. Линеарне комбинације оригиналних варијабли којима се оптимизује дискриминација група у дискриминантној анализи се називају дискриминантне функције. Последњих година постоји тенденција да се назив „дискриминантна анализа” користи за дискриминацију било којег броја група, иако је и назив „каноничка анализа“ још увек у широкој употреби.

Као и код анализе главних компоненти, каноничка анализа израчунава нови скуп варијабли које представљају линеарне комбинације оригиналних варијабли и међусобно су ортогоналне (тј. под правим углом једна у односу на другу), а јединке се потом распоређују у морфолошком простору дефинисаном новим варијаблама. Нове варијабле се називају каноничке варијабле (*canonical variates*), а положај јединки у новом координатном систему одређују скорови каноничких варијабли (ортогоналне пројекције јединки на каноничке варијабле). Међутим, за разлику од анализе главних компоненти која описује положај јединки, каноничка анализа описује положај група јединки у простору, а израчунате каноничке варијабле обухватају максималну разлику између група (у односу на варијансу унутар групе) која остаје необјашњена свим претходним каноничким варијаблама (Слика 26).



Слика 26. Каноничка анализа. А) Дистрибуција седам група јединки у морфолошком простору дефинисаном варијаблима X_1 и X_2 . Црни кружићи представљају центроиду (средњу вредност) групе, а елипсе варијабилност унутра сваке од њих. Б) Приказана је конфигурација средњих вредности група и елипса која одговара пулираној матрици унутар-групних варијанси. Црвене стрелице представљају главне компоненте (PC1 и PC2) пулиране елипсе. Ц) Почетна конфигурација група ротирана тако да главне компоненте (PC1 и PC2) постају осе новог координатног система. Д) Скалирање главних компоненти и трансформација елипсе унутар-групне варијансе у кружни облик. Плаве стрелице представљају главне компоненте средњих вредности група, тј. каноничке варијабле (CV1 и CV2). Извор: модификовано из Campbell и Atchley (1981).

Каноничка анализа започиње анализом главних компоненти пулираних (просечних) унутар-групних варијанси; израчунате главне компоненте чине осе новог координатног система и цео скуп података ће се заротирати, али ће релативни однос између група остати непромењен. Потом се главне компоненте скалирају тако да имају јединичну варијансу унутар група (то се постиже тако што се скорови главних компоненти поделе са квадратним кореном одговарајуће еигенвредности). Скалирањем се унутар-групна варијабилност трансформише из елиптичног облика у кружни облик. Након овог корака релативни положај група је промењен. Ротиране и скалиране осе, које одражавају обрасце варијабилности унутар групе, постају референтне осе координатног система у следећем кораку анализе у коме се међу-групна варијабилност посматра у односу на обрасце унутар-групне варијабилности дефинисане главним компонентама. Примењујући поново анализу главних компоненти, овога пута на центроиде (средње вредности) група у ротираном и скалираном простору, проналази се правац (каноничка варијабла) у коме групе теже да буду најудаљеније (Слика 26). Свака наредна каноничка варијабла представљаће правац који описује највећи могући део разлике између група која није описана претходним каноничким варијаблама. Могући број каноничких варијабли који се може израчунати једнак је броју група мање један.

Дакле, прва каноничка варијабла представља правац у коме су групе најефикасније раздвојене и обухвата највећи део укупне варијансе између група, али то није обавезно и правац у ком се средње вредности група највише разликују. Нови скуп варијабли ће бити израчунат без обзира да ли су разлике између група статистички значајне. Ефикасност појединачних каноничких варијабли у дискриминацији анализираних група јединки може се проценити на основу параметра Wilks' Λ и χ^2 -теста (за процедуру тестирања значајности каноничких варијабли погледати у Zelditch и сар., 2012). Статистички значајне каноничке варијабле указују да се бар неке од анализираних група јединки могу разликовати датим варијаблама.

Пошто се у току каноничке анализе осе новог координатног система рескалирају (тј. пропорционално издужују) у односу на варијабилност унутар групе, положај јединки у координатном систему дефинисаном каноничким осама није само заротиран већ и промењен, па се удаљеност између група у таквом морфолошком простору може разликовати од оне у оригиналном координатном систему.

У алгебарском смислу, каноничка анализа се заснива на утврђивању односа просечне унутар-групне коваријационе матрице (представља одступања јединки од средње вредности своје групе) и међу-групне коваријационе матрице (представља део укупне варијансе који није објашњен унутар-групном матрицом). Како процедура израчунавања каноничких варијабли обухвата инвертовање просечне унутар-групне матрице, величина анализираних узорка мора бити већа од броја варијабли. Величина узорка би требала бити већа од $[(2k - 4) + (G - 1)]$, где је k број варијабли и G број група (Webster и Sheets, 2010). Иако постоје мишљења да би величина узорка требала

бити најмање пет пута већа од броја варијабли, свакако да ће минимална одговарајућа величина узорка зависити од тога колико су групе различите као и од варијабилности у величини група (Strauss, 2010).

Канонички скорови омогућавају да се на графичком приказу морфолошког простора визуелизује раздвајање група. Удаљеност између група у геометријском простору у коме су варијабле некорелисане и једнако скалиране мери се као растојање између центроида (центар, средина) група и назива се Махаланобисова дистанца (D^2). Махаланобисова дистанца се такође може измерити и између појединачне тачке у простору (јединке) и центроиде групе или између две тачке. Каноничке варијабле и Махаланобисове дистанце су основ за процедуру класификовања јединки за које није позната групна припадност. Дакле, на основу морфометријских података за неklasификовану јединку се може утврдити којој од унапред дефинисаних група највероватније припада (под претпоставком процедуре класификације да јединка која се класификује заправо јесте члан једне од група обухваћених анализом). Припадност неklasификоване јединке некој од дефинисаних група може се одредити тако што се израчуна Махаланобисова дистанца јединке од центроиде сваке групе, па се јединка потом додели оној групи са којом има најмању вредност дистанце, односно најближа јој је. За процену вероватноће припадности датом групи (или вероватноће погрешне класификације) може се користити χ^2 дистрибуција (дистрибуција вероватноће на којој се заснива Махаланобисова дистанца), или применити *bootstrap* процедура.

Поузданост са којом морфометријски подаци подржавају унапред дефинисане групе може се утврдити помоћу процедуре унакрсне валидације (*cross-validation*). Ова широко коришћена процедура поновног узорковања за процену статистичких модела започиње поделом анализираног скупа података/јединки на два подскупа; први подскуп јединки је за тренинг (калибрацију) и служи за прилагођавање модела подацима и израчунавање параметара, а други подскуп јединки служи за тестирање, односно за процену перформанси модела. Дакле, на подскуп за тренинг се примени каноничка анализа, а потом се јединке подскупа за тестирање третирају као да је њихова припадност одређеним групама непозната и на основу претходно израчунатих каноничких варијабли и Махаланобисових дистанци јединки од центроида група изврши се њихово распоређивање у дефинисане групе. Ови подскупи се формирају случајним распоређивањем јединки, при чему скуп за тестирање обично садржи мали број јединки (од једне јединке до 10% величине узорка). Читава процедура (формирање подскупова, израчунавање варијабли, додељивања јединки у групе) се поновља више пута, а успешност којом су јединке разврстане у *a priori* дефинисане групе исказује се стопом коректне класификације јединки. Стопа коректне класификације би требало да одражава укупну успешност методе каноничке анализе и може бити користан прилог тестирању разлика између група у просечном облику. Метода унакрсне валидације је посебно прикладна и када је величина узорка мала.

Резултати каноничке анализе (илустровани примером геометријско морфометријске анализе варијабилности облика крила три врсте комараца) садрже

листу генерисаних каноничких варијабли и вредност варијансе за сваку од њих (еигенвредности), проценат укупне варијансе објашњен сваком варијаблом (Табела 8) и скорове каноничких варијабли за сваку јединку за сваку варијаблу (Табела 9).

Табела 8. Резултати каноничке анализе варијабли облика (Прокрустове координате) крила врста *Aedes vexans*, *Anopheles maculipennis* и *Culex pipiens* (Diptera, Culicidae)

Каноничка варијабла (CV)	Варијанса	% варијансе	Кумулативни % варијансе
1	85,6683	81,824	81,824
2	19,0304	18,176	100

Табела 9. Индивидуални скорови каноничких варијабли добијени каноничком анализом варијабли облика (Прокрустове координате) крила врста *Aedes vexans*, *Anopheles maculipennis* и *Culex pipiens* (Diptera, Culicidae). Приказане су вредности скорова за првих десет (од 150) анализираних јединки

Јединка	CV1	CV2
1	7,2358	-4,0981
2	5,3971	-6,1211
3	4,5113	-7,6441
4	7,6018	-6,3766
5	6,3977	-4,6567
6	6,8714	-5,2405
7	6,4899	-4,4183
8	5,7714	-4,9925
9	5,516	-6,6455
10	6,0922	-3,7293

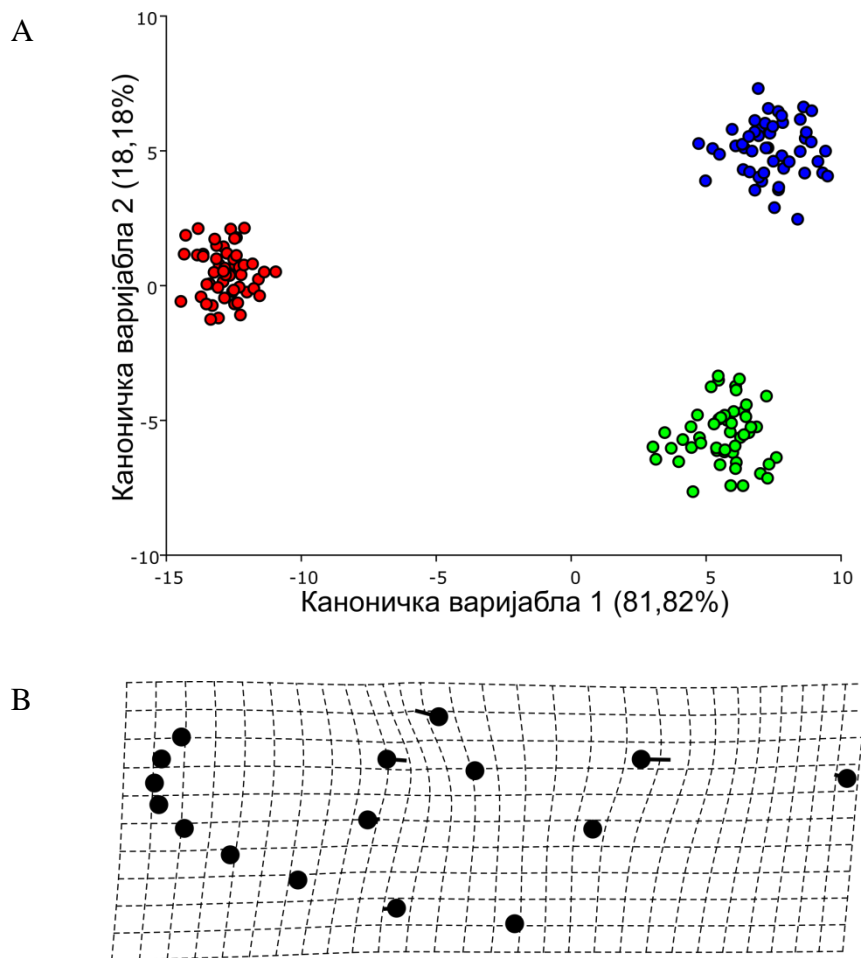
Такође, приказују се и резултати класификације укључујући број успешно (коректно) и погрешно класификованих јединки у односу на *a priori* претпостављену припадност одређеној групи и процену поузданости утврђену унакрсном валидацијом (Табела 10). Дискриминација група у морфолошком простору дефинисаном каноничким осама приказује се у виду дводимензионалног или тродимензионалног графика расејања (Слика 27А). У случају када су анализом обухваћене две групе, графички приказ је у виду хистограма (број каноничких варијабли који се генерише када се анализирају две групе је један, па није могуће приказати графиком расејања). Као и у анализи главних компоненти, множење оригиналних варијабли са коефицијентом каноничких варијабли и сумирање даје серију вектора релативног померања специфичних тачака који илуструју варијабилност облика дуж каноничких оса (Слика 27Б).

Табела 10. Матрица коректне класификације јединки врста *Aedes vexans*, *Anopheles maculipennis* и *Culex pipiens* (Diptera, Culicidae). Редови представљају *a priori* дефинисану припадност, а колоне претпостављену припадност одређеној групи. Бројеви ван дијагонале указују на степен неуспеха класификације. Процент коректне класификације износи 96%; од укупно 151 анализиране јединке, коректно је класификовано 145.

	<i>Aedes vexans</i>	<i>Anopheles maculipennis</i>	<i>Culex pipiens</i>	Укупно
<i>Aedes vexans</i>	47	2	1	50
<i>Anopheles maculipennis</i>	0	49	2	51
<i>Culex pipiens</i>	1	0	49	50
Укупно	48	51	52	151

Када је величина узорка мала или број варијабли велик, потенцијални проблеми који се могу јавити у каноничкој анализи могу се избећи редукцијом димензионалности података. Ово се може постићи применом каноничке анализе на скуп главних компоненти (видети анализу главних компоненти). Други приступ је примена међу-групне анализе главних компоненти (*between-groups Principal component analysis; bgPCA*) која анализира разлике између средина група без обзира на величину или образац унутар-групне варијабилности. Другим речима, ова алтернативна метода представља пројекцију података на главне компоненте средина група (Mitteroecker и Bookstein, 2011). На овај начин се елиминише дисторзија дистанце између средина група до које долази услед рескалирања оса у каноничкој анализи, што омогућава анализу разлика између средњих вредности група без промене њихове међусобне позиције или дистанце у простору. Такође, димензионалност морфолошког простора са

великим бројем варијабли се редукује на број варијабли који не може бити већи од броја група минус један (Zelditch и сар., 2012).



Слика 27. Графички резултати каноничке анализе варијабли облика (Прокрустове координате) крила врста *Aedes vexans*, *Anopheles maculipennis* и *Culex pipiens* (Diptera, Culicidae). А) График расејања приказује дистрибуцију јединки (зелени кружићи - *Ae. vexans*, црвени кружићи - *An. maculipennis*, плави кружићи *C. pipiens*) у морфопростору дефинисаном са прве две каноничке варијабле. Б) Мрежа деформација облика крила дуж прве каноничке варијабле. Положај специфичних тачака у конфигурацији просечног (консензус) облика приказан је црним кружићима, а векторима (црне линије) релативна промена њиховог положаја са повећањем скорa каноничке варијабле.

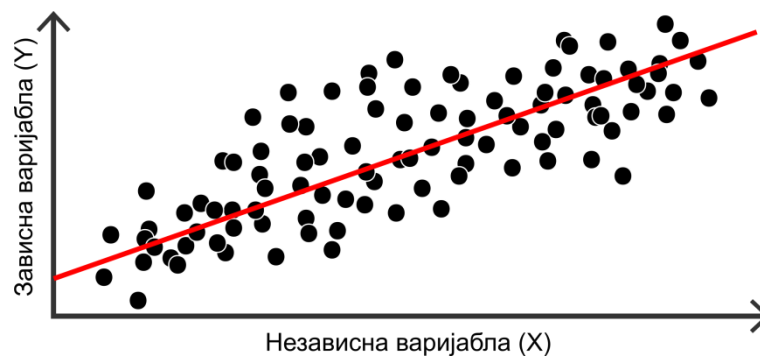
Мултиваријантна анализа варијансе (*Multivariate analysis of variance*; MANOVA) се користи за утврђивање статистичке значајности разлика у облику између група. За разлику од ANOVA анализе, MANOVA подразумева да су морфометријски подаци мултиваријантни, те се може применити на варијаблама облика (Прокрустове координате или на скорови јединки добијени у анализи главних компоненти) пројектованим у тангентни простор или на главним компонентама генерисаним у анализи главних компоненти варијабли облика (Zelditch и сар., 2012). Као што је раније већ речено, ANOVA помоћу F -теста квантификује разлику између средњих вредности група (варијанса између група) у односу на варијансу унутар групе. Слично, MANOVA мери различитост између мултиваријантних центроида (средња вредност) група у односу на просечну унутар-групну коваријациону матрицу. Наиме, разлика између група у облику се изражава Прокрустовом дистанцом између средњих вредности група, а F -тест представља однос фактором објашњене (међу-групне) и необјашњене (унутар-групне) варијансе ових дистанци (Goodall, 1991; Dryden и Mardia, 1998). У пракси, MANOVA израчунава параметар Wilks' Λ који представља удео унутар-групне варијансе (необјашњени део који се назива и резидуал или грешка) у укупној варијанси и може имати вредност од нула до један. Ако је израчуната вредност за Λ мала, то указује да групе имају различите средње вредности (велики део укупне варијансе чини варијанса објашњена груписањем). Потом се Λ вредност конвертује обично у F статистику или χ^2 вредност (мада су доступне и друге апроксимације за трансформисање Λ у алтернативне статистике), а на основу којих се утврђује статистичка значајност разлика било параметријским или пермутационим приступом. Када се MANOVA примени на скуп варијабли које нису корелисане и утврди да постоји статистички значајна разлика између група, онда се могу урадити за сваку варијаблу засебно ANOVA да би се проценила која од варијабли је допринела разликама у групи. Међутим, ако су варијабле које се анализирају у корелацији (као што је то случај са координатама конфигурације специфичних тачака), онда одвојене ANOVA анализе нису статистички независне, те их није оправдано радити. Када је MANOVA примењена за тестирање разлике у средњим вредностима за више од две групе, анализа се може наставити упоређивањем парова група *post hoc* тестом.

Када је варијабилност облика крила три врсте комараца (*Ae. vexans*, *An. maculipennis* и *C. pipiens*) анализирана применом MANOVA, израчунати параметри Wilks' $\Lambda = 0,0006$ и $F_{(56,242)} = 175,73$ су указали да је разлика у облику крила између анализираних врста статистички значајна ($P < 0,001$). Прокрустова дистанца између *Ae. vexans* и *An. maculipennis* износи 0,0947, *Ae. vexans* и *C. pipiens* 0,0835, а између *An. maculipennis* и *C. pipiens* 0,1319.

Регресиона анализа (*Regression analysis*) се примењује за тестирање хипотеза о међусобном односу између варијабли. Тестирање значајности односа између варијабли нам може рећи да ли је зависност између њих слаба, што значи да нам знање о независној варијабли не омогућава да предвидимо зависну варијаблу или је, пак, однос статистички значајан што значи да објашњава значајан део варијабилности зависне варијабле. Регресионом анализом се утврђује облик зависности анализираних варијабли

(линеарна, нелинеарна, вишеструка, директна, инверзна). То се постиже применом одговарајућег регресионог модела (статистички модел који математичким формулама и при одређеним претпоставкама најбоље описује зависност између варијабилности посматраних варијабли), док се постојање и јачина статистичке везе између варијабли одређује корелационом анализом.

Најједноставнији је модел са две варијабле (биваријантни модел), једна зависна и једна независна варијабла, у коме је веза између варијабли одређена линеарном функцијом (тј. једначином праве линије) чији график је у виду праве линије, те се назива и линеарна регресија (Слика 28). Једначина праве линије има облик $Y = a + bX$, где су X и Y независна и зависна варијабла, коефицијент a је одсечак на y оси (место где регресиона линија сече y осу), а коефицијент b представља нагиб регресионе линије. Дакле, регресиони коефицијент b представља просечну промену зависне варијабле за јединично повећање независне варијабле. Значајност регресије одређује се помоћу F -теста (однос варијансе објашњене регресионим моделом и оне која није објашњена моделом, тј. резидуалне). Тако се, на пример, линеарна регресија може применити за утврђивање односа између појединачних главних компоненти и величине, односно да ли величина (независна варијабла) утиче на варијабилност облика дуж осе једне главне компоненте.

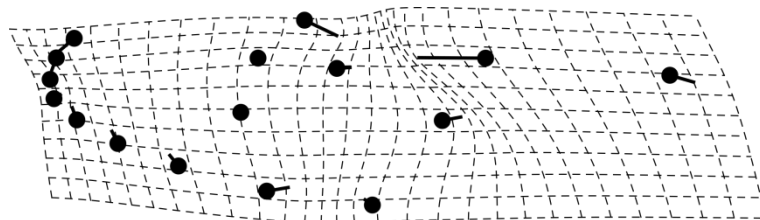


Слика 28. Линеарна регресија са две варијабле. Црне тачке представљају јединице посматрања дистрибуиране у простору дефинисаном независном варијаблом X и зависном варијаблом Y . Црвена линија која пролази кроз скуп података назива се регресиона линија и описује статистичку везу између независне и зависне варијабле.

Модел регресије примењен на мултиваријанте податке назива се мултиваријантна регресија, где се подаци за облик (координате специфичних тачака или скорови главних компоненти) као зависна варијабла регресирају на независну варијаблу (величина, еколошка варијабла, старост и слично). Као и у случају модела са две варијабле, израчунавају се коефицијенти нагиба и одсечка регресионе праве, док се тестирање значајности регресије заснива на израчунавању параметра Wilks' Λ и његовој апроксимацији тј. конверзији у F вредност.

Поред значаја да се варијабилност величине и облика анализирају и сагледају засебно, често нас интересује и како ова два својства коварирају, односно у којој мери варијабилност у величини утиче на варијабилност облика анализiranог објекта. Овај ефекат се назива алометрија и односи се на промену облика која је у вези са разликама у величини објекта. Пошто алометрија може да објасни велики и статистички значајан део морфолошке варијабилности, анализа облика се може извести и са варијаблама облика које не садрже алометријску компоненту варијабилности. (Напомињемо да се у кораку рескалирања током процедуре Прокрустове суперимпозиције конфигурације специфичних тачака скалирају за изометријску промену величине, а алометријска варијабилност и даље остаје укључена у варијабле облика.) Варијабле облика кориговане за алометрију се добијају мултиваријантном регресијом варијабли облика и величине, којом се, као што је раније речено, варијабилност зависне варијабле раздвоји на две компоненте (варијабилност предвиђена независном варијаблом и резидуална варијабилност). Израчунати резидуали управо представљају нове варијабле облика које нису у корелацији са независном варијаблом (величином), односно варијабле облика из којих је уклоњен ефекат величине.

У примеру три врсте комараца, мултиваријантном регресијом варијабли облика на центроидну величину крила је утврђено да удео алометрије у укупној варијабилности облика износи 19,65% и да је статистички значајан (Wilks' $\Lambda = 0,1916$; $F_{(28,122)} = 18,390$; $P < 0,001$). Ефекат величине на облик може се графички приказати помоћу мрежа деформација облика (Слика 29).



Слика 29. Мрежа деформација облика крила са променом центроидне величине крила врста *Aedes vexans*, *Anopheles maculipennis* и *Culex pipiens* (Diptera, Culicidae).

Анализа парцијалних најмањих квадрата (*Partial least squares analysis*; PLS) испитује однос између два или више сетова (блокова) варијабли прикупљених на истим објектима, при чему за разлику од регресионе анализе не претпоставља да су скупови података међусобно зависни (Rohlf и Corti, 2000; Bookstein и сар., 2003). Наиме, овом методом се проучавају обрасци коварирања између скупова варијабли које могу чинити варијабле облика (барем један од блокова садржи податке о облику), али такође и друге варијабле. Тако се, на пример, може проучавати коварирање облика и варијабли животне средине (климатске, географске, биотичке и слично), функције, или понашања (видети на пример, Rohlf и Corti, 2000; Zelditch и сар., 2012; Deitloff и сар., 2013; Yazdi и сар., 2015). Такође, метода се може користити за анализу коваријанси између два или

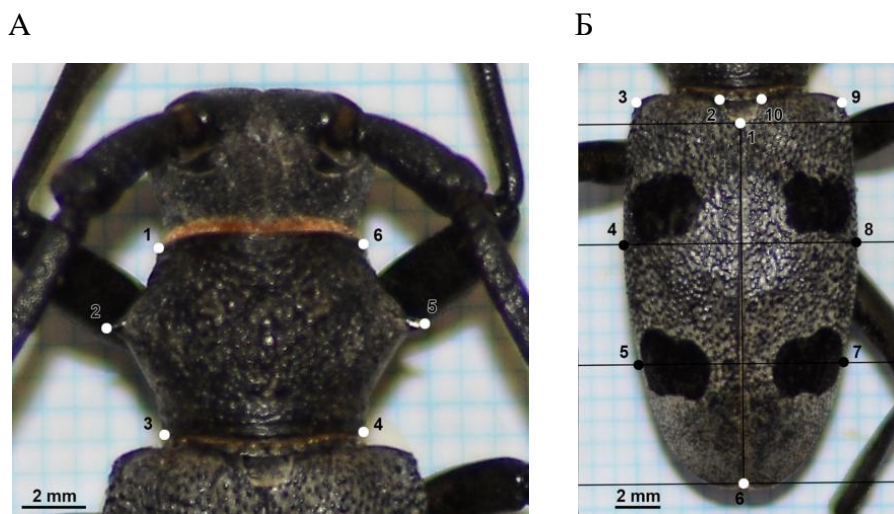
више блокова варијабли облика, било да се ради о варијаблама облика различитих анатомских структура једног организма или скуповима варијабли облика делова једне анатомске структуре, а тиме и за процену морфолошке интеграције и модуларности анализираних структура (Klingenberg, 2009).

Морфолошка интеграција и модуларност су уско повезани концепти еволуционе развојне биологије (*Evolutionary Developmental Biology; evo-devo*) који се односе на обрасце и процесе интеракције и независности особина. Фенотипска својства организама не варирају независно, већ су она међусобно интегрисана интеракцијама различите јачине. Дакле, морфолошка интеграција описује степен усаглашености варијабилности морфолошких структура или организма у целини (на пример, различитих морфолошких структура у организму, или делова у оквиру одређене структуре). При томе, поједине структуре унутар интегрисане целине, услед различитих развојних, функцијских или еволуционих ограничења, могу показивати већи степен међусобне интеграције и чинити тако релативно независне модуле. Један од задатака морфометријских истраживања јесте да се утврди да ли одређена структура представља јединствену интегрисану јединицу или се, пак, састоји од неколико различитих модула, те да их и идентификује. Модули су делови биолошких система који представљају комплексе особина чврсто интегрисаних многим и често врло јаким интеракцијама између њих, а који су релативно независни од других модула; интеракције између различитих модула су слабије или их је мање (Klingenberg, 2009, 2010; Armbuster и сар., 2014). Дакле, очекује се да ће оне особине које су функционално интегрисане, или на чију морфогенезу утичу заједнички развојни фактори бити међусобно повезане у већој мери (и чинити један модул), него са особинама које не припадају истој функционалном или развојном комплексу (Lawing и Polly, 2009). Модуларност је пронађена у многим (али не и свим) проучаваним структурама; тако је, на пример, утврђен одређен степен модуларности за доњу вилицу и лобању сисара, док јака морфолошка интеграција (тј. одсуство модуларне структуре) карактерише крила мува и цврчака (Klingenberg, 2009, 2010).

Метода парцијалних најмањих квадрата, попут анализе главних компоненти и каноничке анализе, редукује димензионалност података тако што генерише нове варијабле (линеарне комбинације оригиналних варијабли обично називане PLS осе или сингуларне осе) које описују коваријансу између скупова варијабли, при чему су поређане од пара који објашњава максималну коваријансу до пара који објашњава најмању. Коваријанса између два блока се квантификује помоћу R^2 коефицијента који представља мултиваријантну верзију стандардне униваријантне корелације (Escoufier, 1973; Klingenberg, 2009). Вредност R^2 коефицијента може бити у распону од 0 (нема коварирања) до 1 (потпуно коварирање), а статистичка значајност се може тестирати пермутационим тестом (за више детаља видети Klingenberg, 2009). За квантификовање

јачине коваријације могу се користити и Z скорови (Adams и Collyer, 2016), чија се статистичка значајност такође процењује пермутационим тестом.

Као пример једне од примена анализе парцијалних најмањих квадрата можемо навести анализу морфолошке варијабилности облика две анатомске структуре врсте *Morimus asper* (Coleoptera, Cerambycidae) (Gojković и сар., 2022). Наиме, применом геометријске морфометрије анализирана је варијабилност облика пронотума (горња плоча проторакса) и елитри (очврсла предња крила), те степен коварирања ове две структуре засебно код женки и мужјака (Слика 30). Резултати анализе (обично) обухватају листу PLS оса и величину коваријансе (сингуларна вредност и % укупне коваријансе) објашњене сваком осом уз пратећу P вредност добијену тестирањем значајности, корелације између скорова за сваки пар оса и пратећу P вредност (Табела 11), те RV коефицијент. Анализом је утврђено статистички значајно коварирање две структуре код оба пола (женке: $RV = 0,271$, $P < 0,05$; мужјаци: $RV = 0,384$, $P < 0,01$) (Gojković и сар., 2022). Такође, мрежама деформације су приказане промене облика дуж PLS оса (Слика 31А), док се повезаност два блока може приказати помоћу графика расејања индивидуалних скорова PLS оса једног блока у односу на други блок (Слика 31Б).

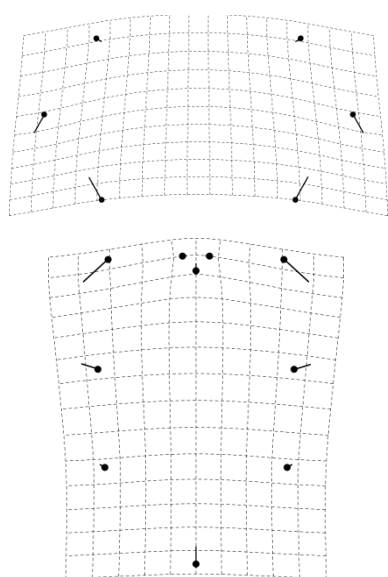


Слика 30. Конфигурације специфичних тачака (бели кружићи) и полутачака (црни кружићи) за морфометријску анализу *Morimus asper* (Coleoptera, Cerambycidae): А) пронотум; Б) елитре. Извор: модификовано из Gojković и сар. (2022).

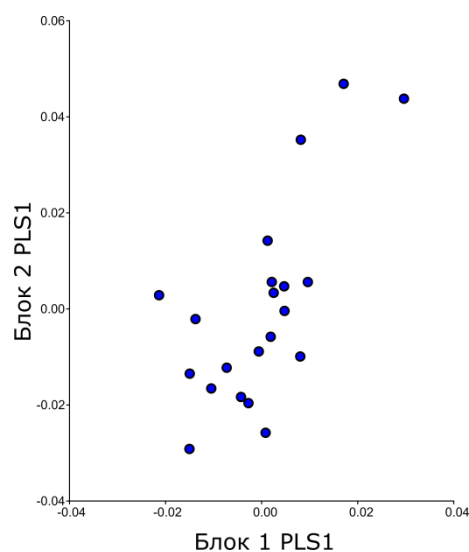
Табела 11. Резултати анализе парцијалних најмањих квадрата варијабли облика пронотума и елитри женки и мужјака врсте *Morimus asper* (Coleoptera, Cerambycidae)

	PLS оса	Сингуларна вредност	<i>P</i>	% укупне коваријације	Корелација	<i>P</i>
Женке	PLS1	0,00012581	0,027	88,874	0,62776	0,074
	PLS2	0,00003882	0,452	8,461	0,50167	0,200
	PLS3	0,00001673	0,704	1,572	0,41123	0,228
	PLS4	0,00001395	0,099	1,093	0,31536	0,170
Мужјаци	PLS1	0,00019119	0,0009	97,032	0,70582	0,003
	PLS2	0,00002904	0,7976	2,239	0,36931	0,487
	PLS3	0,00001568	0,6834	0,653	0,27596	0,569
	PLS4	0,00000537	0,7819	0,077	0,22081	0,426

А

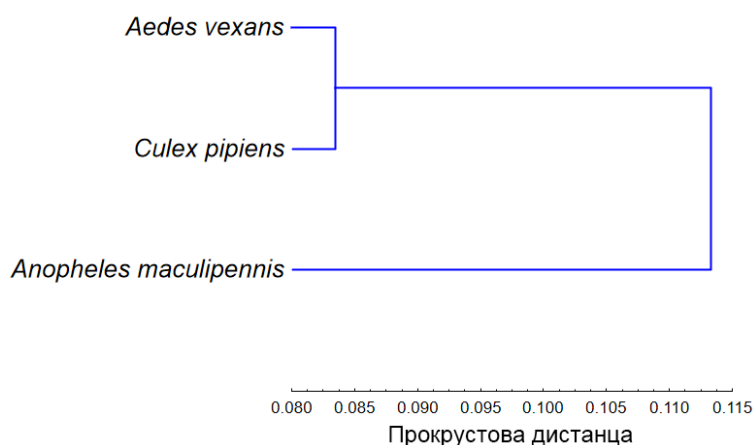


Б



Слика 31. Графички резултати анализе парцијалних најмањих квадрата варијабли облика пронотума и елитри женки и мужјака врсте *Morimus asper* (Coleoptera, Cerambycidae). А) Мрежа деформација облика пронотума и елитри дуж прве PLS осе. Б) График расејања индивидуалних скорова једног пара PLS оса (PLS1) два блока (блок 1 - варијабле облика пронотума; блок 2 - варијабле облика елитри).

Кластер анализа представља статистичку технику која се користе за класификовање елемената скупа података у релативне, дискретне групе који се називају кластери, тако да су елементи у оквиру исте групе сличнији једни другима (у односу на анализирано својство) него онима у другим кластерима. Дакле, циљ анализе је да се пронађе оптимално груписање на основу вредности измерених својстава, што омогућава сагледавање структуре анализираних података. Методе кластер анализе могу бити хијерархијске и нехијерархијске методе (Sneath и Sokal, 1973). Међу бројним доступним алгоритмима у биолошким истраживањима се највише користе методе из групе узастопних, агломеративних, хијерархијских, непреклапајућих кластер метода (Sneath и Sokal, 1973). Ова група обухвата више метода груписања од којих су најпознатије и најчешће коришћене UPGMA (*Unweighted pair group method with arithmetic averages*), метода најближег суседа (*Nearest neighbour clustering, Single linkage*), метода најудаљенијег суседа (*Complete linkage*) и метода по Ward-у. Заједничко за све њих је да анализа започиње формирањем матрице дистанце између елемената анализираних скупа. Постоји више начина за израчунавање коефицијената сличности/различитости и формирање матрице дистанце; у морфометријским истраживањима обично се израчунава Еуклидова дистанца, мада и матрице Махаланобисове дистанце и Прокрустовете дистанце добијене током каноничке анализе/дискриминантне анализе такође могу бити основа за кластер анализу. На основу вредности дистанце врши се постепено (у више узастопних корака) спајање елемената у веће групе (кластере) тако што се групишу прво елементи између којих је вредност дистанце најмања, а потом се у наредним корацима процеса претходно формираном кластеру придружују елементи/кластери следећи по вредности дистанце. Разлика између метода је у начину како се утврђује дистанца на којој се групишу кластери (на пример, у UPGMA методи кластери се спајају на основу просечне удаљености између свих чланова две групе, а у методи најближег суседа на основу најмањег растојања између две групе; за више детаља о методама видети у Sneath и Sokal, 1973). Резултати хијерархијске анализе кластера обично се представљају у форми дендрограма, али како се у морфометријској анализи кластери формирају на основу сличности у морфометријским подацима (на пример, облику) добијено дрво се назива и фенограм. UPGMA дрво конструисано на основу матрице Прокрустових дистанци између три врсте комараца приказано је на Слици 32.



Слика 32. UPGMA дрво конструисано на основу матрице Прокрустових дистанци у облику крила врста *Aedes vexans*, *Anopheles maculipennis* и *Culex pipiens* (Diptera, Culicidae).

Дакле, хијерархијске методе кластер анализе формирају кластере корак по корак док се сви елементи скупа не нађу на дендограму, па се потом утврђује број кластера. За разлику од њих, нехијерархијске методе подразумевају да истраживач унапред дефинише број кластера (обично на основу хипотезе коју тестира) у које ће бити разврстани елементи анализираног скупа. Метода *k*-средина (*k-means method*) је често коришћена нехијерархијска метода.

Еуклидова дистанца представља најкраћи размак између две тачке у линеарном (тангентном) простору.

Прокрустова дистанца представља метричку меру разлике облика; то је Еуклидова дистанца између две конфигурације Прокрустових координата: квадратни корен суме квадрата удаљености између кореспондирајућих специфичних тачака две конфигурације након суперимпозиције. Такође, односи се и на растојање између две конфигурације специфичних тачака у Кендаловом простору облика.

Махаланобисова дистанца је растојање између две тачке у мултиваријантном простору; квадрат удаљености између средина две групе подељен са пулираном коваријационом матрицом.

Наведене су дефиниције према: Mitteroecker и Gunz (2009) и Zelditch и сар. (2012).

Будући да се анализа облика заснива на сложеним математичким процедурама, развој специјализованих алгоритама и статистичких програма за извођење геометријско морфометријске анализе (укључујући програме за прикупљање дводимензионалних и тродимензионалних морфометријских података, Прокрустову суперимпозицију конфигурација специфичних тачака и полутачака, униваријантну и мултиваријантну анализу, визуелизацију варијабилности облика) је био неминован и неопходан. Данас су доступни бројни програми за морфометријску анализу који се могу бесплатно користити у научне и наставне сврхе (за више информација и преузимање бесплатно доступних програма видети на <https://sbmorphometrics.org/>). Међу њима налазе се програми који су креирани за извођење одређене операције или анализе, а који су део интегрисаног низа, односно серије програма, попут *tps* серије (о програмима у оквиру *tps* серије видети раније) и IMP серије. У оквиру серије сви програми користе исти формат датотеке и креирани су тако да буду међусобно компатибилни, односно раде заједно.

- Интегрисани морфометријски пакет (*Integrated Morphometrics Package*, IMP; Sheets, 2014) представља скуп софтверских алата намењених за приказ и анализу геометријских морфометријских података заснованих на конфигурацији специфичних тачака. Пакет не садржи алгоритама за прикупљање морфометријских података, па се подаци у IMP програме могу учитати у виду *.tps датотеке. IMP серија обухвата програме за извођење бројних морфометријских анализа, укључујући суперимпозицију конфигурација специфичних тачака и полутачака, анализу главних компоненти, каноничку анализу, MANOVA, регресиону анализу, анализу парцијалних најмањих квадрата, као и за визуелизацију промена облика. Иако су програми примарно намењени за анализу дводимензионалних морфометријских података, IMP садржи и неколико програма за анализу тродимензионалних података.

Насупрот програмским серијама, постоје и свеобухватни програми који дају могућност извођења многих анализа у оквиру само једног програма, попут MorphoJ (Klingenberg, 2011) и PAST (Hammer и сар., 2001).

- MorphoJ (Klingenberg, 2011) је интегрисани програмски пакет за геометријско морфометријску анализу дводимензионалних и тродимензионалних података конфигурација специфичних тачака. Пошто не подржава прикупљање координата специфичних тачака, не садржи методе за анализу података крива и контура, нити за трансформисање криве/контуре у низ полутачака, подаци прикупљени другим програмима се читавају у MorphoJ (на пример, у *.tps формату). Програм укључује бројне функције за анализу и визуелизацију промена облика, укључујући Прокрустову суперимпозицију података (са и без асиметрије), анализу главних компоненти, каноничку анализу/дискриминантну анализу, регресиону анализу, анализу парцијалних најмањих квадрата, тестирање хипотезе модуларности, тестирање филогенетског сигнала облика и реконструкцију предачког облика

(упутство за коришћење програма видети у Прилогу 4).

- PAST (Hammer, 2001) је свеобухватан статистички пакет који није специјализован искључиво за морфометријску анализу, већ обухвата бројне функције за манипулацију и графичко приказивање података, униваријантну и мултиваријантну статистичку анализу, еколошку анализу, временске серије и просторну анализу, морфометрију и стратиграфију. Доступне опције обезбеђују извођење стандардних морфометријских анализа на подацима конфигурација специфичних тачака (учитавање података је у виду *.tps датотеке). Такође, програм има опцију и за прикупљане морфометријских података, укључујући координате специфичних тачака, линеарне дистанце и углове.
- У групу свеобухватних софтвера за морфометријску анализу може се убројати и програмски језик R (R Core Team, 2018) са бројним доступним кодовима и пакетима, попут пакета *geomorph*, *shapes*, *Morpho* и *Momocs* (Adams и Otárola-Castillo, 2013; Bonhomme и сар., 2014; Schlager, 2017; Dryden, 2021; Adams и сар., 2022). R је програмски језик и софтверско окружење за статистичку анализу и графичко приказивање података. Помоћу функција и пакета, у оквиру јединственог окружења може се извршити прикупљање података, широк спектар статистичких анализа и графички приказати резултати. Такође, корисник може развити и модификовати сопствене или програмиране функције.

Иако многи програми обављају стандардне морфометријске анализе, ниједан од њих не ради све што раде и други. Зато је често неопходна употреба неколико различитих програма.

6. Анализа морфолошког диверзитета - примери

Данас је квантитативна анализа облика морфолошких целина и/или њених делова неизоставна у бројним областима биолошких истраживања. Тако, на пример, подаци добијени анализом варијабилности морфометријских карактеристика организама дају значајан допринос разумевању еволуционих процеса и механизма који учествују у стварању, одржавању и обликовању варијабилности популација. Историјски гледано, сагледавање биолошке разноврсности, таксономија и систематика организама су утемељени на описивању облика и варијацији облика. Међутим, након вишедеценијске (више од 250 година) доминантне примене компаративне морфологије, у последњој деценији XX века нове методе и подаци, углавном молекуларни, преузимају примат (Radial и сар., 2010). Истраживања су показала да информативност различитих типова маркера, података и метода у проучавању биолошког диверзитета често може бити ограничена, те сугерисала на неопходност интегративног приступа. И заиста, бројне студије засноване на комбинованој употреби морфометријских података и молекуларних, еколошких, географских и/или онтогенетских података су показале да ови подаци нису међусобно конкурентни већ комплементарни, те да је интегративни

приступ веома ефикасан, не само у раздвајању и формалном опису номиналних врста (интегративна таксономија), већ и у проучавању структурираности популација (просторна и временска дистрибуција диверзитета).

Практична примена морфометријске анализе у сагледавању дистрибуције фенотипског диверзитета и квантификовању интраспецијске и интерспецијске варијабилности илустрована је на примерима датим у наставку. Предмет приказаних студија био је интегративно истраживање засновано на молекуларним маркерима и фенотипским карактерима, при чему је фенотипска варијабилност анализирана применом традиционалне и геометријско морфометријске методе.

6.1. Просторна дистрибуција морфолошког диверзитета врсте *Telestes dabar* (fam. Leuciscidae)

Врста *Telestes dabar*, народног назива дабарска гаовица, је стено-ендемична слатководна риба из породице Leuciscidae која насељава водотоке Дабарског крашког поља (Опачица, Прибитул, Сушки поток и Вријека) у Босни и Херцеговини (Bogutskaya и сар., 2012; Francuski и сар., 2019). Јединке пронађене у водотоцима Опачица и Вријека у Дабарском пољу су првобитно биле идентификоване као врста *T. metohiensis*. Накнадном таксономском ревизијом, а на основу генетичких и морфолошких података, утврђено је да оне ипак представљају засебну, криптичну врсту, блиско сродну са *T. metohiensis* и именоване су као *T. dabar* (Bogutskaya и сар., 2012). Иако се ради о врсти са веома ограниченим распрострањењем која је свакодневно изложена бројним стресним факторима (попут рекултивације земљишта, загађења водотокова, интродукције алохтоних и инвазивних врста, изловљавања, климатских промена и др.) који потенцијално могу угрозити њен опстанак, статус угрожености према Међународној унији за заштиту природе (*International Union for Conservation of Nature*; IUCN) још увек није утврђен због нејасног таксономског положаја, те недостатка података о молекуларном и фенотипском диверзитету врсте, њеној екологији и биологији. Наиме, прецизна и недвосмислена идентификација врсте, утврђивање њене дистрибуције и сагледавање популационе структуре су први важни кораци приликом дефинисања и заштите еволуционо значајних јединица конзервације (*Evolutionarily Significant Unit*; ESU). У том погледу, Francuski и сар. (2019) су спровели истраживање у коме су применом молекуларних маркера митохондријске ДНК (мтДНК) утврдили границе врсте и еволуциону независност врста *T. dabar* и *T. metohiensis*, а потом и анализирали популациону структуру врсте *T. dabar*.

У поменутом истраживању (Francuski и сар., 2019) популациона структурираност и просторна дистрибуција морфолошког диверзитета врсте *T. dabar* анализирана је применом традиционалне и геометријско морфометријске анализе. Узорак је обухватао 70 јединки врсте *T. dabar* сакупљених на четири локалитета у Дабарском пољу (локалитети: Опачица - 16 јединки, Прибитул - 14, Сушки поток - 21,

Вријека - 19). Све јединке су фотографисане у стандардизованим условима (објектив апарата је био причвршћен за статив, постављен под правим углом у односу на бочну површину рибе и на истој удаљености од објекта код свих фотографисаних јединки). На основу дигиталних фотографија у програму tpsDig (Rohlf, 2015) на телу сваке јединке је измерено (у мм) девет линеарних дистанци (Слика 1) и постављено 15 специфичних тачака (Слика 3). Како би се уклонио утицај величине, измерене вредности линеарних дистанци су трансформисане на два начина: први, израчунавањем процентуалног удела дате мере у некој од мера тела (SL, HL или CPL) и други, помоћу формуле $M_{adj} = M \times (L_s/L_o)^b$ (Elliott и сар., 1995). Потом је статистичка анализа линеарних мера изведена за оба скупа трансформисаних варијабли. Конфигурације специфичних тачака јединки су подвргнуте процедури исправљања искривљеног положаја тела (операција *unbend specimens* у програму tpsUtil; Rohlf, 2015) узрокованог небиолошким факторима (на пример, услед чувања материјала). Специфична тачка број 15 позиционирана је на бочној линији на једнакој удаљености од тачака 6 и 11, а након кориговања искривљености тела ова тачка је искључена из даље анализе. За конфигурације специфичних тачака је потом урађена Прокрстова суперимпозиција, а добијене центроидна величина (варијабла величине) и Прокрстова координате (варијабле облика) су даље коришћене у статистичкој анализи варијабилности величине и облика. Све морфометријске анализе су изведене у програмима MorphoJ 1.06d (Klingenberg, 2011) и PAST 3.20 (Hammer и сар., 2001).

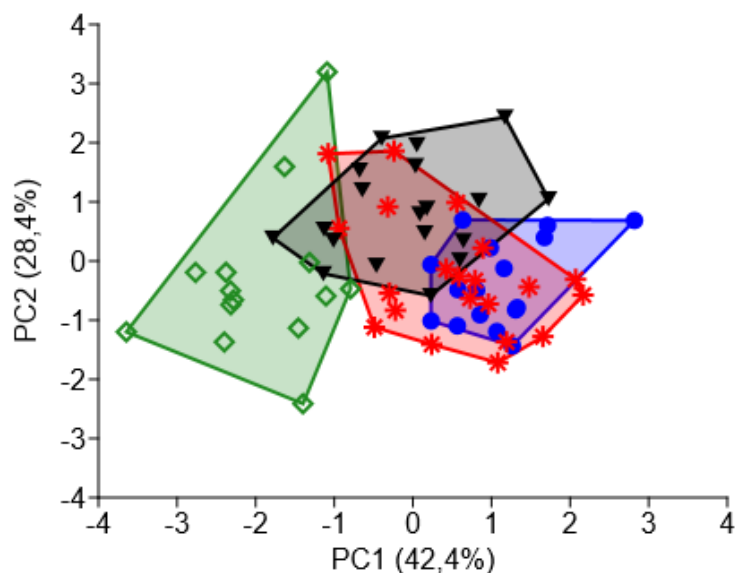
Резултати традиционалне морфометрије. Стандардна дужина риба у узорку била је у распону од 50,55 до 105,89 мм. Дескриптивна статистика линеарних мера четири популације је показала да се за сваку појединачну варијаблу распон варирања (најмања - највећа вредност) преклапа између популација, док се аритметичке средине разликују (Табела 12). Значајност уочене разлике између аритметичких средина четири популације тестирана је за сваку варијаблу понаособ применом анализе варијансе (ANOVA). Како ANOVA подразумева да анализирана варијабла има нормалну дистрибуцију, тестом нормалности (Shapiro-Wilk тест) је проверено да ли све варијабле испуњавају овај предуслов. За четири варијабле које су показале одступање од нормалне дистрибуције тестирање је изведено Kruskal-Wallis тестом (непараметарска ANOVA). Статистички значана разлика ($P < 0,05$) између аритметичких средина четири популације утврђена је за већину анализираних варијабли (Табела 12).

Табела 12. Средња вредност и стандарна девијација ($\bar{x} \pm SD$) линеарних морфометријских података за четири популације *Telestes dabar* (Leuciscidae). Значајност разлике аритметичких средина тестирана је применом ANOVA и Tukey теста (статистички значајни парови популација наведени су у загради; $P < 0,05$). Извор: модификовано из Francuski и сар. (2019)

Варијабла	Опачица $\bar{x} \pm SD$	Прибитул $\bar{x} \pm SD$	Сушки поток $\bar{x} \pm SD$	Вријека $\bar{x} \pm SD$	ANOVA
SL (мм) ^a	82,37±11,87	74,36±12,26	80,99±9,02	77,69±9,85	нс
OL(%SL)	6,01±0,45	6,17±0,46	5,86±0,35	5,97±0,35	нс
OL(%HL)	24,85±1,52	25,47±1,79	24,50±1,77	25,69±1,37	нс
SNL(%SL)	6,12±0,36	6,09±0,51	5,74±0,64	5,47±0,48	** (O/B)
SNL(%HL)	25,31±1,29	25,11±1,44	23,91±1,66	23,49±1,34	*** (O/C,O/B, П/В)
PHL(%HL)	52,20±1,69	52,14±1,44	53,91±1,74	52,69±1,44	**
HL(%SL)	24,16±0,63	24,23±1,11	23,98±1,24	23,27±1,07	*
BD(%SL) ^a	20,37±0,90	21,06±2,12	20,46±1,02	21,14±1,00	нс
CPD(%SL)	9,17±0,41	9,54±0,49	9,36±0,50	9,81±0,41	*** (O/B)
CPD(%CPL)	44,77±2,99	46,47±3,97	45,58±4,08	46,30±3,41	нс
CPL(%SL) ^a	20,54±1,11	20,63±1,40	20,62±1,23	21,26±1,33	нс
OL _{adj}	4,92±0,22	4,56±0,18	4,74±0,26	4,72±0,29	** (O/П)
SNL _{adj}	5,03±0,27	4,53±0,38	4,66±0,52	4,37±0,38	*** (O/П,O/B)
PHL _{adj}	10,39±0,34	9,39±0,43	10,46±0,48	9,73±0,38	*** (O/П,O/B, П/СП, C/B)
HL _{adj}	19,88±0,44	17,99±0,75	19,42±1,00	18,49±0,68	*** (O/П,O/B, П/С, C/B)
BD _{adj}	16,77±0,73	15,68±1,51	16,57±0,83	16,95±0,85	** (П/В)
CPD _{adj}	7,56±0,33	7,08±0,34	7,58±0,40	7,78±0,39	*** (O/П,П/В, П/С)
CPL _{adj}	16,93±0,86	15,27±0,76	16,72±0,94	16,91±0,83	*** (O/П,П/С, П/В)

^a - линеарне мере тестиране Kruskal-Wallis тестом; нс - није статистички значајно; * $P < 0,05$; ** $P < 0,01$; *** $P < 0,001$; O - Опачица; B - Вријека; П - Прибитул; C - Сушки поток

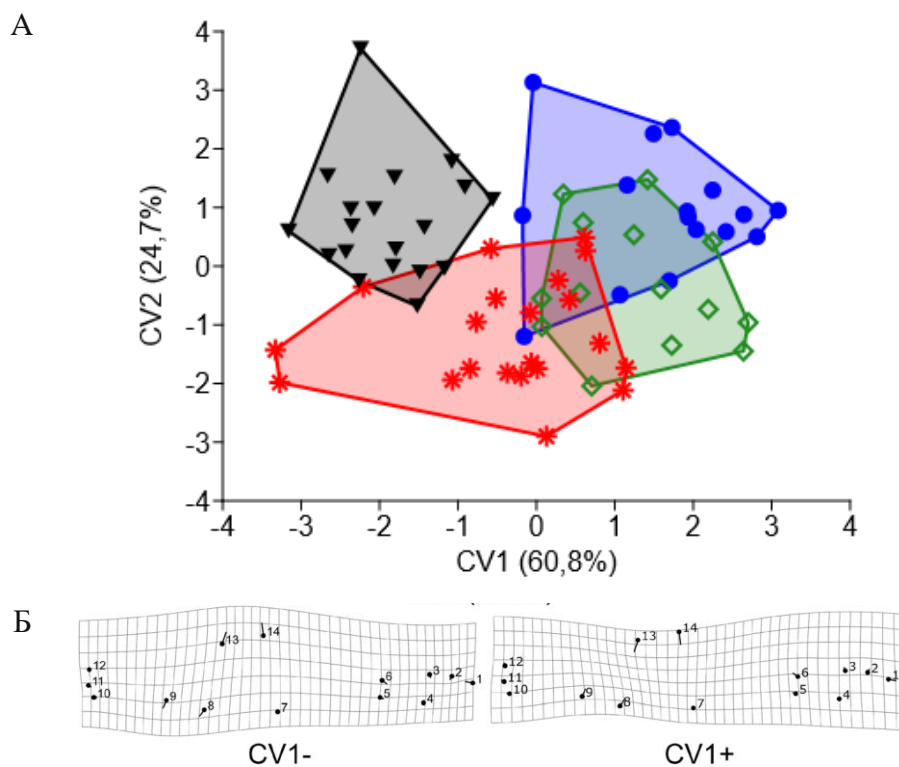
Поређење парова *post hoc* Tukey тестом је показало да се од шест тестираних парова популација број статистички значајних поређења разликује између варијабли (1 до 4 пара популација) (Табела 12). Интересантно је уочити да, иако скуп линеарних дистанци трансформисан на два начина не даје конзистентне резултате, статистички значајна разлика између парова популација Опачица/Прибитул и Опачица/Вријека је утврђена на основу већине анализираних мера. Поред униваријантне статистичке анализе, на скуп од седам варијабли трансформисаних према формули коју је дао Elliott и сар. (1995) примењена је и анализа главних компоненти. Резултати ове анализе су показали да више од 70% укупне варијабилности се може објаснити са прве две главне компоненте (PC1 42,2%, PC2 28,4%) и да су највећи допринос имале варијабле HL_{adj} (0,63) и BD_{adj} (0,72). На графику расејања се уочава да се у морфолошком простору јединке популације Прибитул дуж прве главне компоненте (PC1) издвајају у односу на остале популације, док се дуж друге осе (PC2) раздвајају јединке популација Опачица и Вријека (Слика 33) (Francuski и сар., 2019).



Слика 33. График расејања добијен анализом главних компоненти линеарних морфометријских података (подаци су трансформисани према Elliott и сар., 1995) јединки четири популације *Telestes dabar* (Leuciscidae). Опачица - плава боја, Прибитул - зелена боја, Сушки поток - црвена боја, Вријека - црна боја. Извор: модификовано из Francuski и сар. (2019).

Резултати геометријске морфометрије. ANOVA центроидне величине тела анализираних риба указала је на одсуство статистички значајне разлике ($F_{(3,66)} = 1,83$; $P = 0,15$) између популација *T. dabar*. Насупрот величини, каноничка анализа и мултиваријантна анализа (MANOVA) варијабли облика су показале статистички значајне разлике у облику тела риба ($Wilks' \Lambda = 0,097$; $F_{(72,129)} = 2,12$; $P < 0,001$). Поређењем парова популација, статистички значајна Прокрустова дистанца утврђена је

између парова популација Вријека/Опачица (Прокрстова дистанца = 0,0156; $P < 0,01$), Вријека/Прибитул (Прокрстова дистанца = 0,0178; $P < 0,001$) и Вријека/Сушки поток (Прокрстова дистанца = 0,0120; $P < 0,01$). Такође, на графику расејања који приказује морфопростор дефинисан са прве две каноничке осе (CV1 и CV2) које обухватају 60,8% и 24,7% укупне варијабилности облика уочава се да се популација из Вријеке јасно диференцира у односу на преостале три популације (Слика 34А). Мреже деформација указују да су промене облика дуж прве каноничке осе резултат промена у положају специфичних тачака 8, 9, 13 и 14, односно у висини тела риба (Слика 34Б) (Francuski и сар., 2019).



Слика 34. График расејања (А) и мреже деформација (Б) добијени каноничком анализом варијабилности облика (Прокрстова координата) тела риба четири популација *Telestes dabar* (Leuciscidae). Опачица - плава боја, Прибитул - зелена боја, Сушки поток - црвена боја, Вријека - црна боја. Положај специфичних тачака у конфигурацији просечног (консензус) облика приказан је црним кружићима, а векторима (црне линије) релативна промена њиховог положаја са повећањем скова каноничке варијабле. Извор: модификовано из Francuski и сар. (2019).

Закључак. Резултати морфометријске анализе указују на морфолошку хетерогеност врсте *T. dabar* што је вероватно резултат утицаја различитих срединских фактора у стаништима ових популација. Наиме, познато је да је фенотипска

варијабилност, посебно облик тела, тесно повезан са варијаблама животне средине (рН земљишта, брзина воде, температура и другим), разликама у исхрани и притисцима од стране предатора (Collin и Fumagalli, 2011). У исто време, анализом молекуларних података утврђена је и интраспецијска генетичка варијабилност, као и одсуство молекуларне диференцијације између популација. Дакле, у раду утврђена морфолошка (и молекуларна) варијабилност сугерише на постојање фенотипске пластичности и адаптивног потенцијала ове недавно описане врсте. Поред тога, утврђена генетичка хомогеност четири анализирани популације је потврдила да дистрибуција врсте *T. dabar* у Дабарском пољу није ограничена само на Опачицу и Вријеку, те да је присутна и у Прибитулу и Сушком потоку (Francuski и сар., 2019).

Интересантно је приметити да постоји одређено неслагање у погледу популационе диференцијације квантификоване линеарном и геометријском морфометријском анализом. Ова неконзистентност (одступање) између добијених резултата је врло вероватно условљена чињеницом да различити морфометријски приступи (линеарна и геометријска морфометрија) и типови морфометријских карактера (линеарне дистанце и конфигурације специфичних тачака) нису једнако осетљиви за детектовање фенотипске варијабилности (Francuski и сар., 2019).

6.2. Интерспецијска и интраспецијска варијабилност врста *Ochlerotatus caspius* и *O. dorsalis* (Diptera, Culicidae)

Комарци (fam. Culicidae) представљају групу инсеката која поред значајне улоге у функционисању екосистема (чине важне карике у воденим и копненим ланцима исхране) има и велики епидемиолошки и медицински значај с обзиром на чињеницу да су бројне врсте комараца вектори изазивача разних обољења људи и домаћих животиња. За креирање ефикасних програма контроле комараца заснованих на правилном мониторингу врста од епидемиолошког значаја, утврђивање граница врста између блиско сродних врста и њихова тачна идентификација је неопходна. Квантификовање морфолошког и генетичког диверзитета унутар и између популација важно је за одређивање таксономског статуса организама.

Због свог епидемиолошког значаја и неразјашњеног таксономског статуса криптични таксони *Ochlerotatus caspius* и *O. dorsalis* су били предмет компаративне анализе засноване на интегративној употреби морфолошких и молекуларних података (Francuski и сар., 2021). Том приликом проучена је варијабилност секвенци гена за субјединицу I цитохром *c* оксидазе митохондријске ДНК (*COI* мтДНК) и унутрашње маргине 2 гена за рибозомалну ДНК једарног генома (*ITS2* рДНК), те меристичких и морфометријских карактеристика симпатричких популација две врсте у циљу процене валидности одабраних молекуларних и морфолошких маркера за идентификацију врста и сагледавања њихове интраспецијске варијабилности и филогеографске структуре.

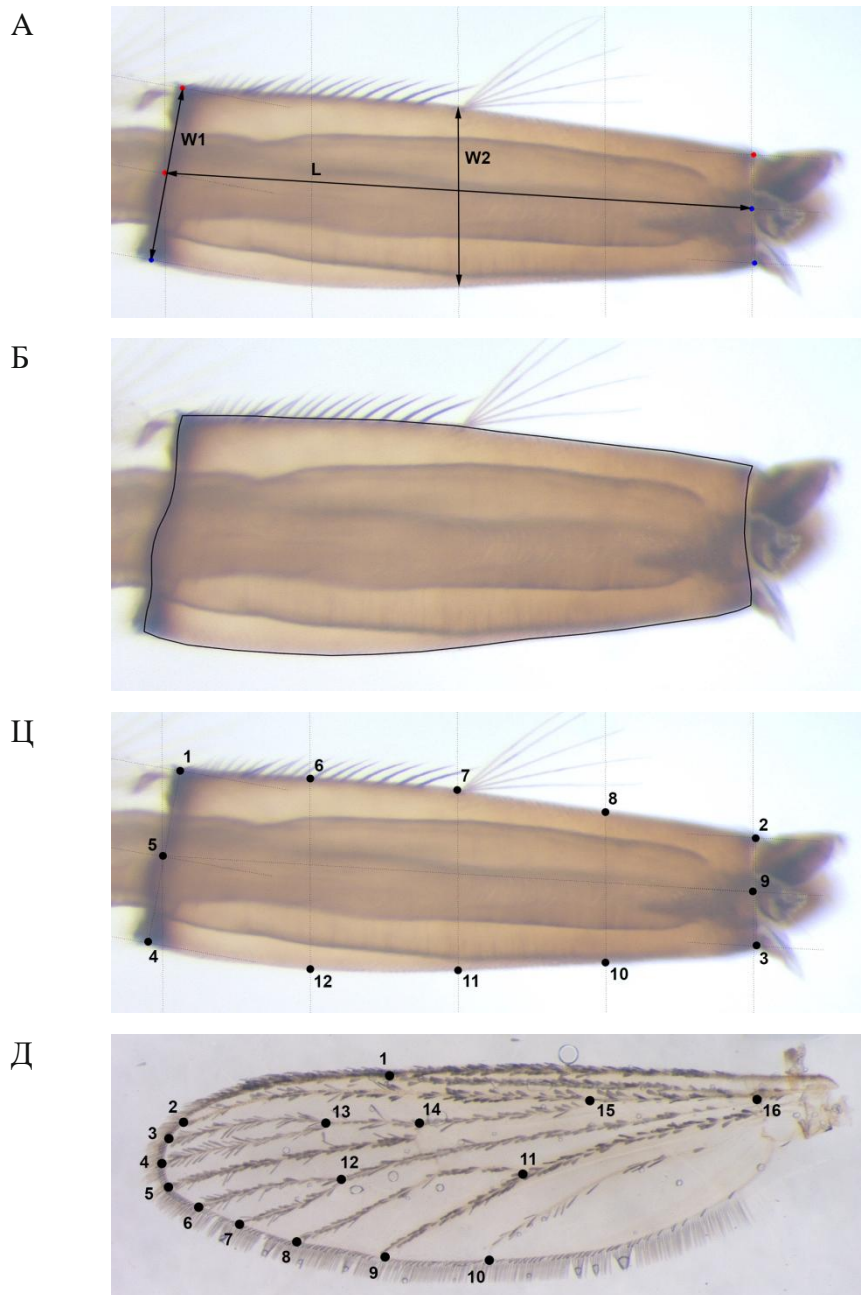
За анализу фенотипске варијабилности прикупљени су меристички и морфометријски подаци за 113 ларви четвртог ступња (73 јединке *O. caspius* и 40 јединки *O. dorsalis*) и 120 адултих јединки (97 јединки *O. caspius* и 23 јединки *O. dorsalis*) две врсте (Francuski и сар., 2021). Наиме, прикупљене су фотографије респираторне цевчице (сифон) ларви и левог крила адулта помоћу дигиталне камере прикључене на стереомикроскоп у стандардизованим условима (исто увећање). Потом су на основу фотографија прикупљени и меристички и морфометријски подаци (Francuski и сар., 2021). Утврђен је број зубића пектена и број длака у чуперку сифона. Помоћу програма tpsDig (Rohlf, 2015) измерене су дужина сифона (L), ширина сифона на најширем делу и ширина на средини сифона (W1, W2) (Слика 35А), као и обим и површина сифона на основу исцртане контуре (Слика 35Б). За сваку јединку на основу вредности дужине и ширине сифона израчуната су два сифонална индекса ($SI1 = L/W1$; $SI2 = L/W2$) и индекс ширине ($WI = W1/W2$), а на основу вредности површине, обима и дужине сифона и индекси обима ($PI = P/L$) и површине ($AI = A/L$). Такође, истим програмом постављене су и специфичне тачке и полутачке на сифону (Слика 35Ц) и крилу (Слика 35Д). Положај полутачака дефинисан је помоћу мреже једнако удаљених линија постављених програмом IMP серије MakeFan8 (Sheets, 2014).

Статистичка анализа меристичких (број зубића пектена и број длака у чуперку) и морфометријских (израчунати индекси, центроидна величина и варијабле облика) података изршена је у програмима Statistica 13 (TISCO Software Inc., 2018) и PAST 3.24 (Hammer и сар., 2001).

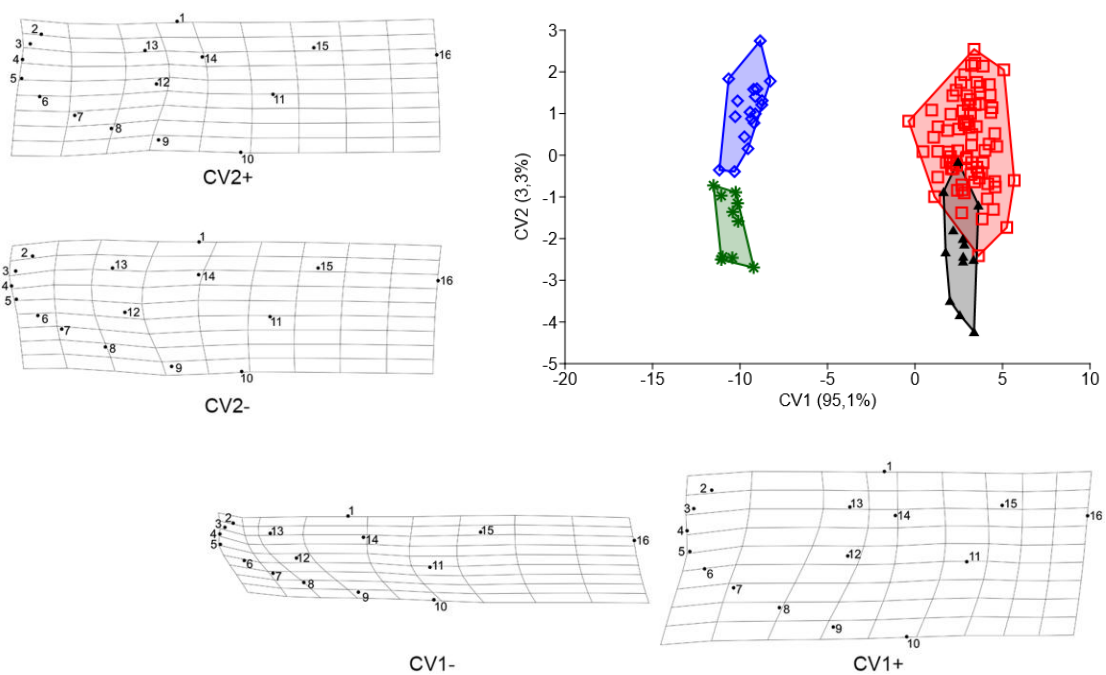
Резултати морфометријске анализе карактера ларви. Двофакторска ANOVA (фактори: врста и пол) меристичких података, индекса линеарних мера и центроидне величине сифона открила је статистички значајне разлике између анализираних врста на основу броја зубића пектена ($F_{(1,109)} = 20,76$; $P < 0,001$), индекса $SI1$ ($F_{(1,109)} = 4,99$; $P < 0,05$), $SI2$ ($F_{(1,109)} = 6,40$; $P < 0,01$), PI ($F_{(1,109)} = 5,66$, $P < 0,05$) и центроидне величине ($F_{(1,109)} = 5,71$; $P < 0,05$). Такође, двофакторска MANOVA варијабле облика је указала на статистички значајне разлике између две врсте у облику сифона (Wilks' $\Lambda = 0,616$; $F_{(20,92)} = 2,87$, $P < 0,001$). Насупрот јасној интерспецијској диференцијацији, ни један од анализираних морфолошких карактера није указао на присуство полног диморфизма на ларвалном ступњу развића (Francuski и сар., 2021).

Резултати морфометријске анализе карактера адулта. Значајне разлике између врста детектовала је и двофакторска ANOVA/MANOVA центроидне величине ($F_{(1,116)} = 4,90$; $P < 0,05$) и облика крила (Wilks' $\Lambda = 0,448$; $F_{(28,91)} = 4,01$, $P < 0,001$). Међутим, анализа варијабилности облика крила је указала и на јасну диференцијацију између женки и мужјака (тј. полни диморфизам; Wilks' $\Lambda = 0,033$; $F_{(28,91)} = 95,98$, $P < 0,001$). Каноничка анализа облика крила издвојила је три каноничке осе од којих је прва (CV1) обухватала 95,1% укупне варијабилности облика; графички приказ индивидуалних каноничких скорова показује раздвајање јединки различитих полова дуж CV1 осе, док се јединке две врсте раздвајају дуж друге осе (CV2) (Слика 36). Визуелизацијом промена облика помоћу мрежа деформација уочава се да женке карактеришу шира, краћа и заобљенија крила, док је разлика у облику крила између

врста сконцентрисана у централном делу крила услед промене у положају специфичних тачака 1, 9, 12 и 13 (Слика 36) (Francuski и сар., 2021).



Слика 35. Морфометријски подаци прикупљени за анализу фенотипске варијабилности врста *Ochlerotatus caspius* и *O. dorsalis* (Diptera, Culicidae). А) Линеарне мере дужине (L) и ширине (W1, W2) сифона. Б) Контура сифона. Ц) Положај специфичних тачака (тачке 1 - 4) и полугачака (тачке 5 - 12) на сифону. Д) Положај специфичних тачака на крилу. Извор: модификовано из Francuski и сар. (2021).



Слика 36. График расејања и мреже деформација добијени каноничком анализом варијабли облика крила врста *Ochlerotatus caspius* и *O. dorsalis* (Diptera, Culicidae). Плава боја - *O. caspius* мужјаци, црвена боја - *O. caspius* женке, зелена боја - *O. dorsalis* мужјаци, црна боја - *O. dorsalis* женке. Промене облика су увећане 3×. Положај специфичних тачака на мрежи одговара положају тачака приказаном на Слици 35Д. Извор: модификовано из Francuski и сар. (2021).

Закључак. У раду Francuski и сар. (2021) анализом интерспецијске и интраспецијске морфолошке варијабилности врста *O. caspius* и *O. dorsalis* је потврђен дијагностички значај у раздвајању врста морфолошких карактера ларви који су се традиционално користили у таксономији комараца (број зубића пектена, сифонални индекс), док је дијагностичка вредност броја грана у чуперку сифона била ниска. Поред тога, центроидна величина и облик респираторне цевчице и крила су се показали као поуздани карактери за успешно развајање анализираних врста. Утврђена интерспецијска диференцијација на основу морфолошких карактеристика подржана је и резултатима анализе молекуларне варијабилности (на основу секвенци *COI* мтДНК гена). Сагледавање дистрибуције морфолошке варијабилности унутар врста указало је на одсуство полног диморфизма на преадултном ступњу развића, а значајне разлике између полова код адулта. Конкретно, облик крила је једини од анализираних карактера који се значајно разликовао између женки и мужјака. Наиме, током онтогенетског развића комарци мењају животну средину (ларве се развијају у воденој средини, а адулти у ваздушној), па коришћење ресурса различитих станишта током сваке од фаза развоја може условити и различите обрасце морфолошке диференцијације унутар и између популација (Francuski и сар., 2021).

ДЕО II
КВАНТИФИКОВАЊЕ
ГЕНЕТИЧКОГ ДИВЕРЗИТЕТА

7. Предеона генетика

Сматрајући генетичку варијабилност за најосновнији ниво биолошког диверзитета и предуслов за варијабилност популација, врста, и екосистема, научници различитих биолошких дисциплина, укључујући еволуциону биологију, екологију и конзервациону биологију, настоје да сагледају и разумеју факторе и процесе који обликују генетичку варијабилност у природи (Primack, 2014; Balkenhol и сар., 2016). Наиме, генетички диверзитет је пресудан за разлике у адаптивној вредности (тј. разлике у преживљавању и репродукцији јединки), опстанак популација и способност врста да се адаптирају на промене у животној средини (Frankham и сар., 2010).

Генетичка варијабилност подразумева разлике у наследном материјалу на одређеној позицији у геному између јединки унутар популације. Унутар генома постоје локације које могу бити под дејством природне селекције (називају се адаптивни локуси), што значи да је у овим локусима количина генетичке варијабилности и учесталост одређених алела у популацији директно под утицајем срединских услова у станишту. Наиме, адаптивни локуси директно или индиректно утичу на фенотип и адаптивну вредност јединке, а предео и фактори спољашње средине утичу на учесталост алела деловањем селекције; на пример, природна селекција у датим условима средине фаворизује алеле који су повољни (тј. повећавају адаптивну вредност јединке), чиме се повећава њихова фреквенција. Са друге стране, у геному постоје и делови који нису под утицајем селекције (ове локације се називају адаптивно неутрални локуси) и учесталости алела у њима су примарно под утицајем протока гена и генетичког дрифта (Hedrick, 2011; Waits и Storfer, 2016). Управо због чињенице да на генетичку варијабилност популације или врсте утичу четири главна еволуциона процеса - мутација, природна селекција, проток гена и генетички дрифт, за разумевање ових процеса на популационом и геномском нивоу од кључног значаја је анализа генетичког диверзитета унутар врсте.

У сагледавању и разумевању генетичке варијабилности важна су два њена аспекта: 1) генетички диверзитет који представља количину генетичке варијабилности утврђену унутар јединке, популације или географског простора; 2) генетичка структура популације која се односи на дистрибуцију генетичке варијабилности између јединки, популација или области. Процењом ове две главне компоненте генетичке варијабилности бави се популациона генетика. Међутим, на количину и дистрибуцију генетичке варијабилности велики утицај имају и одлике географског предела (*landscape*); тако, на пример, од карактеристика предела зависи које његове делове одређена врста може да насели (што утиче на просторну дистрибуцију генетичке варијабилности), колико јединки живи на некој локацији (што утиче на количину генетичке варијабилности у популацијама), као и степен кретања између јединки на различитим локацијама (што утиче на проток гена између популација) (Waits и Storfer, 2016). Како би утврдили утицај хетерогености средине на проток гена и генетичку варијабилност, научници су почели да комбинују принципе и методе популационе генетике и предеоне екологије, чиме је започео развој нове области истраживања

назване предеона генетика (*landscape genetics*) (Manel и сар., 2003; Storfer и сар., 2007). Дакле, предеона генетика интегрише популациону генетику, предеону екологију и просторне аналитичке технике у циљу експлицитног квантификовања ефекта композиције, конфигурације и/или квалитета матрице предела на микроеволуционе процесе попут протока гена, генетичког дрифта и селекције, а користећи адаптивно неутралне и адаптивне генетичке локусе (Storfer и сар., 2007; Balkenhol и сар., 2016).

Предео (*landscape*) представља систем еколошких образаца и процеса који су у интеракцији на било којој (просторној) скали и може се сматрати подручјем које је просторно хетерогено у односу на најмање један фактор у просторном оквиру одговарајућем за односе образац - процес за дати фактор (Balkenhol и сар., 2016).

Предеона екологија (*landscape ecology*) проучава интеракције између просторних образаца и еколошких процеса; бави се конфигурацијом и просторним оквиром (скалом) у ком се дешавају еколошки процеси и настају одређени обрасци. Теорија и методологија предеоне екологије великим делом је усмерена на то да окарактерише хетерогеност предела, утврди одакле она потиче, како се мења током времена и зашто је важна за еколошке процесе (Manel и сар., 2003; Cushman и сар., 2016).

Дизајнирање и спровођење истраживања предеоне генетике започиње дефинисањем циљева истраживања. У већини студија фокус је на разумевању протока гена у смислу квантификовања ограничења протока гена услед, на пример, постојања баријера (изолација баријером) или географске удаљености (изолација дистанцом). Такође, циљ истраживања може бити усмерен и на разумевање селекције и локалне адаптације организама. Након постављања циљева, неопходно је одабрати одговарајућу просторну и временску скалу (локални или просторно шири, регионални оквир; краћи или дужи временски оквир), начин и стратегију узорковања података (број и просторна дистрибуција прикупљених јединки и карактеристика предела или средине) и генетичке маркере који ће бити анализирани (Hall и Weissing, 2014). Наиме, важно је имати на уму да на изборе направљене током сваког корака дизајнирања и спровођења истраживања предеоне генетике утичу одлуке из претходних корака. Избор просторне и временске скале истраживања ће зависити од постављених циљева истраживања, односно процеса (проток гена или селекција) који се квантификује, али и животне историје и демографије објекта истраживања (дисперзија, узраст, стратегија укрштања, генерацијско време, дистрибуција, понашање, захтеви за стаништем). Тако, на пример, у истраживањима организама који имају велике дисперзивне способности, неопходно је прикупљање података у ширем просторном оквиру (скали). Такође, и одабир генетичког маркера којим ће бити квантификована генетичка варијабилност мора бити одговарајући у односу на постављене циљеве и дизајн истраживања. Важну улогу у развоју предеоне генетике имала је појава и доступност економичних (исплативих) и брзих техника за генотипизацију јединки чиме је знатно олакшана процена генетичког диверзитета унутар популација. Велика количина доступних генетичких података

условила је и развој нових метода за њихову анализу, а које су понекад рачунски врло сложене и захтевне. Ове методе су имплементирани у бројне софтверске пакете и програме који садрже како стандардне (основне), тако и напредне и софистициране анализе. Данас је међу истраживачима опште прихваћен став да правилна анализа било ког сета генетичких података подразумева употребу више различитих приступа, при чему се започиње утврђивањем основних карактеристика података, а потом следе различите специјализоване анализе, често имплементирани у неколико програма.

Дакле, студије предеоне генетике подразумевају прикупљање и анализу генетичких и просторних података кроз три аналитичка корака (Balkenhol и сар., 2016): 1) мерење генетичке варијабилности - ослања се на популационо-генетички приступ и подразумева опис генетичке композиције јединки или популација; 2) квантификовање хетерогености предела - задатак је предеоне екологије и односи се на утврђивање композиције, конфигурације и/или квалитета матрице предела; и 3) статистичко повезивање хетерогености предела и генетичке варијабилности. Генетички и просторни (предеони и средински) подаци се обично прво анализирају одвојено да би се генерисале генетичке и просторне мере које служе као улазни подаци за анализу предеоне генетике, мада се подаци могу користити и у необрађеном облику (сирови подаци), попут генотипова или вредности климатских варијабли. Статистичким методама које интегришу генетичке и просторне податке испитује се однос између генетичких мера заснованих на јединкама или популацијама и просторно експлицитног предела или животне средине (Hall и Beissinger, 2014).

8. Молекуларни подаци

Генетички маркер представља било коју биолошку карактеристику (фенотипска особина, протеин, ген или препознатљива секвенца ДНК) која се наслеђује са родитеља на потомство према Менделовом правилу раздвајања (сегрегације) алела и пружа информацију о генотипу јединке (Kadirvel и сар., 2015). Приликом дизајнирања истраживања, одабир одговарајућег генетичког маркера у односу на постављени истраживачки задатак зависиће од карактеристика самих маркера, доступних информација о анализираном таксону (на пример, о биологији врсте, развијеним протоколима за одређене маркере за дату врсту и слично) и карактеристика генома (на пример, начин наслеђивања, стопа мутације различитих генома и различитих делова унутар одређеног генома, маркери специфични за одређени геном, разлике у перформансама маркера у зависности од типа генома). Својства које одликују одређени генетички маркер поузданим и добрим маркером укључују: 1) полиморфност, јер је степен полиморфизма одраз варијабилности; 2) присуство више алела (генетички полиморфизам) гена датог маркера, високо корелираних са мерљивим варирањем у експресији дате особине; 3) кодоминантно наслеђивање, јер тада хетерозиготни генотип истовремено испољава својства оба родитеља, те је могуће разликовати хомозиготе

међусобно, као и хомозиготе од хетерозигота; 4) одсуство епистатичке интеракције са другим локусима приликом испољавања анализираних својстава; 5) одсуство утицаја срединских фактора, стадијума развића и типа ткива на испољавање особине; 6) равномерна дистрибуција у геному (тј. да нису груписани у кластере у одређеним регионима) и бројност; 7) једноставност и брзина анализе, висока поновљивост и ниска цена (Agarwal и сар., 2008; Chesnokov и сар., 2020). Ипак, генетички маркер који задовољава све наведене критеријуме (“идеалан” маркер) још увек није доступан истраживачима.

Генетички маркери се карактеришу јединственим својствима од којих и зависи њихова употребљивост у решавању различитих истраживачких питања. Наиме, генетички маркери се међусобно могу разликовати у односу на ефекат који селекција има на њих (неутрални или адаптивни), начин наслеђивања (уни- или бипарентално), стопу мутације (ниска, умерена или висока), или њихову примењивост у закључивању у одређеном просторном (мали или велики) и временском (кратак или дуги) оквиру (Hall и Weissing, 2014). Тако, адаптивно неутрални локуси пошто немају утицај (или је он мали) на адаптивну вредност организма одражавају образац протока гена и генетичког дрифта у проучаваном пределу, док, са друге стране, адаптивни локуси, утичући директно или индиректно на фенотип и адаптивну вредност организма, рефлектују образац селекције. Дакле, употреба адаптивно неутралних генетичких података је погодна за сагледавање како предео утиче (или је утицао) на генетичку повезаност популација, те предвиђање како ће се будуће промене предела одразити на ову повезаност и генетичку варијабилност, док нам подаци адаптивних локуса могу указати како средински фактори утичу на географску дистрибуцију генетичке варијабилности (Waits и Storfer, 2016). Начин наслеђивања различитих маркера има утицаја на њихову ефективну величину популације (N_e) и сагледавање образаца протока гена. На пример, маркери једарног генома (нДНК) се наслеђују бипарентално (генетички материјал се наслеђује од оба родитеља, са изузетком полних хромозома), тако да они осликавају образац протока гена од оба родитеља. Са друге стране, маркери генома митохондрија (мтДНК) и хлоропласта (цпДНК) се наслеђују унипарентално (генетички материјал се наслеђује од једног родитеља, обично од мајке) због чега је ефективна величина популације ових локуса четири пута мања у односу на једарне локусе ($1/4N_e$), а ефекат генетичког дрифта на њих већи (Hall и Weissing, 2014). На употребљивост одређених генетичких маркера утиче и њихова стопа мутације; маркери са већом стопом мутације се користе за квантификовање генетичке варијабилности у просторно и временски мањем оквиру (на пример, проучавање блиско сродних популација), док су они са нижим стопама погоднији за истраживања на већим просторним и дужим временским скалама. Наиме, избор одговарајуће просторне/временске скале је изузетно важан корак приликом осмишљавања истраживања и стратегије прикупљања узорака, а зависи од постављених циљева истраживања и карактеристика животне историје и демографије анализираних таксона (дисперзивна способност, узраст, стратегија парења, генерацијско време, дистрибуција, понашање, заступљеност, захтеви у погледу станишта) (Hall и Weissing, 2014).

Појмом генетички маркери обухваћени су маркери различитих категорија: морфолошки маркери (детектовани на нивоу фенотипа организма), цитолошки маркери (детектовани на нивоу хромозома), биохемијски маркери (различити протеини, укључујући ензиме и метаболите) и молекуларни маркери (детектовани на нивоу нуклеинских киселина) (Kumar, 1999; Kadirvel и сар., 2015).

Морфолошки маркери (често се називају и фенотипски маркери, иако се појам фенотип односи на све спољашње и унутрашње структуре и функције организма које се могу описати и проучавати морфо-анатомским и физиолошким методама) су обично визуелно уочљива манифестација активности гена у његовом нормалном или мутираном облику. Дакле, морфолошке особине су већином резултат експресије гена који се могу користити као генетички маркери. То их чини једноставним и лако доступним за употребу, без потребе за применом специјализованих и/или скувих биохемијских и молекуларних техника. Ови маркери обухватају читав низ гена који контролишу фенотипске карактеристике и били су први маркери коришћени за процену диверзитета унутар и између популација (Mirzaei, 2021). Ипак, недостаци који могу да лимитирају употребу морфолошких маркера су њихова ограниченост у броју (у односу на ДНК маркере), експресија на коју утичу фактори средине, епистатичке и плејотропне интеракције, те чињеница да у случају потпуне доминансе на локусима није могуће разликовати јединке хетерозиготног генотипа од оних са хомозиготним (Chesnokov и сар., 2020)¹.

Цитолошки маркери представљају јединствене карактеристике хромозома у погледу морфологије, структуре и броја хромозома. Другим речима, ови маркери се односе на варијабилност у броју, величини и облику хромозома, обрасцима трака (боја, ширина, редослед и положај трака на хромозому указују на разлику у дистрибуцији еухроматина и хетерохроматина) и динамици хромозома током мејозе (Kadirvel и сар., 2015). Карактеристици их мањи степен полиморфизма и ограниченост у броју.

Биохемијски маркери су протеини, попут ензима, структурних, резервних, регулаторних (хормони), заштитних (антитела) и других група протеина. У популационо генетичким студијама од протеинских маркера се најчешће користе ензими или алоензими кодирани алелним варијантама структурних гена. Алозими су ензими детерминисани различитим алелима истог генског локуса, док су изозими производ различитих гена. Наиме, мутација у ДНК ланцу која доводи до замене аминокиселине (несинонимна мутација) у протеину резултира формирањем структурних варијанти одређеног ензима које имају различите молекулске тежине и електрофоретску покретљивост, а исту каталитичку активност или функцију. Алелна варијабилност ензимских локуса се детектује применом методе гел електрофорезе која подразумева раздвајање протеинских молекула у електричном пољу и потом специфично бојење гела на коме се раздвајање дешавало како би се визуелизовала места до којих су одређени ензимски молекули мигрирали. Интерпретацијом добијеног зимограма (образац трака раздвојених ензима) утврђује се индиректно број локуса и

¹ О методама анализе варијабилности морфолошких карактеристика видети у Делу I Квантификовање морфолошког диверзитета.

алела који детерминишу дати ензим, те генотип јединке. Разлике у релативним фреквенцијама алела се користе за квантификовање генетичке варијабилности и разликовање генетичких јединица на нивоу популација, врста и виших таксономских категорија. Ензими су кодоминантни маркери који имају велику поновљивост. Једноставни су за употребу, а лабораторијске анализе су брзе и са релативно ниским трошковима. Међутим, употреба ових маркера је ограничена чињеницом да је број полиморфних локуса који се могу анализирати, као и ниво полиморфизма у локусима често мали. Наиме, број локуса који се могу визуелизовати техникама хистохемијског бојења је већи од 300, али то и даље представља само мали део целокупног генома (Murphy и сар., 1996). Поред тога, сматра се да су ензими селективно неутрални маркери, а пошто их карактерише нижа стопа мутације у односу на друге генетичке маркере погодни су за квантификовање генетичког диверзитета у студијама дизајнираним да обухвате већу просторну/временску скалу (Hall и Beissinger, 2014). Протеински маркери се често убрајају у молекуларне маркере (јер представљају варијанте ензима, а ензими су молекули), мада се у суштини ради о фенотипским маркерима.

8.1. Молекуларни маркери

Молекуларни маркери обухватају бројну групу маркера за утврђивање полиморфизма ДНК и РНК молекула. Последњих деценија, употреба ДНК маркера постала је рутинска захваљујући бром развоју молекуларних техника и алата за прецизну, брзу и финансијски приступачну процену генетичке варијабилност. До данас развијене технологије молекуларних маркера се заснивају на неколико различитих приступа и метода у детектовању маркера, укључујући хибридизацију, ланчану реакцију полимеразе (*Polymerase Chain Reaction*; PCR) и секвенцирање. Маркери засновани на хибридизацији представљају прву генерацију молекуларних маркера; визуелизација ДНК профила се постиже хибридизацијом фрагмента претходно исечене испитиване ДНК са обележеним (радиоактивно или нерадиоактивно) ДНК пробама познатог порекла и секвенце. Са открићем PCR технике развија се нова генерација маркера који подразумевају амплификацију (умножавање) одређене ДНК секвенце или локуса уз помоћ специфичних или произвољно одабраних олигонуклеотидних секвенци (*primer*; прајмер) и термостабилног ензима ДНК полимеразе. Дакле, захваљујући PCR методи, из мале почетне количине ДНК узорка се може добити велики број копија одређене ДНК секвенце, а гел електрофорезом PCR продуката детектовати варијабилност у дужини умножених секвенци ДНК анализираних узорака. Међутим, треба имати на уму да се на основу обрасца трака на гелу могу детектовати само оне разлике у дужини амплификованих фрагмента које су резултат промена у више нуклеотида. Једино секвенцирањем амплификованих фрагмента могу се открити све разлике између узорака. Наиме, сматра се да су у погледу генетичког диверзитета најпрецизнији и најинформативнији молекуларни

подаци добијени секвенцирањем познатих региона у геному (Karp и сар., 1996). Упоредни преглед основних карактеристика маркера који се најчешће користе у анализи генетичког диверзитета приказан је у Табели 13.

Полиморфизам дужине рестрикционих фрагмената (*Restriction Fragment Length Polymorphisms*; RFLP) се квантификује третирањем геномске ДНК рестрикционим ензимима који секу ДНК ланац на тачно одређеним местима (рестрикциона места; на пример, ензим EcoRI сече ланац на месту дефинисаном палиндромском секвенцом GAATTC између нуклеотида G и A). На тај начин настали фрагменти ДНК (рестрикциони фрагменти) се помоћу методе електрофорезе раздвајају на гелу, а потом се техникама блотовања (*southern blotting*) и хибридизације са обележеним ДНК секвенцама (RFLP пробе) врши идентификација RFLP трака и алела. Мутација која се јавља у секвенци рестрикционог места условљава разлике у дужини рестрикционих фрагмената, па се тако и јединствени образац RFLP трака одређеног генотипа разликује у односу на други генотип. RFLP маркери су специфични за локус, углавном кодоминантни и високо поновљиви (Williams и сар., 1990; Kadirvel и сар., 2015). Поред тога што су процедуре блотинга и хибридизације дуготрајне, а захтевају и велику стручност истраживача и дизајниране обележене RFLP пробе, неопходно је обезбедити и довољну количину висококвалитетне (недеградоване) ДНК (Karp и сар., 1996).

Случајно умножена полиморфна ДНК (*Random Amplified Polymorphic DNA*; RAPD) су фрагменти ДНК добијени PCR амплификацијом насумичних сегмената геномске ДНК са прајмером произвољне нуклеотидне секвенце. Наиме, прајмери се везују за геномску ДНК на оним местима где је редослед нуклеотида комплементаран секвенци прајмера и уколико се прајмери везани за оба ланца ДНК налазе на одређеној удаљености и 3' крајевима окренути један према другом доћи ће до амплификације датог сегмента ДНК. Мутација у ДНК ланцу на месту које је претходно било комплементарно секвенци прајмера резултираће изостанком PCR реакције. Такође, мутација може креирати и ново место за везивање прајмера, а структурне промене попут делеције и инсерције утицати на повећавање или смањивање удаљеност места где су везани прајмери, па настају RAPD маркери различите дужине. Присуство и/или одсуство продукта PCR реакције (амплификованих сегмената ДНК) се детектује њиховим раздвајањем методом електрофорезе и визуелизацијом на гелу. Дакле, полиморфизам RAPD маркера се на гелу огледа у различитим обрасцима присуства/одсуства трака, а као последица разлике у секвенци ДНК анализираних јединки. RAPD маркери су углавном доминантни, нису претерано скупи и захтевни за употребу, али су слабо поновљиви (Williams и сар., 1990; Kadirvel и сар., 2015).

Полиморфизам дужине умножених фрагмената (*Amplified Fragment Length Polymorphism*; AFLP) представља доминантне маркере који се генеришу тако што се прво рестрикционим ензимима исече геномска ДНК, а потом применом PCR методе са специфично дизајнираним прајмерима изводи селективна амплификација рестрикционих фрагмената. PCR продукти се раздвајају и визуелизују на гелу електрофорезом. Полиморфизам AFLP маркера се јавља када мутација, инсерција, делеција или дупликација узрокује промену у секвенци рестрикционог места или места

везивања прајмера, што се у обрасцу трака на гелу може уочити као одсуство или присуство траке, те присуство трака различите величине (а тиме и различите електрофоретске мобилности). Велика покривеност генома и висока поновљивост су предности AFLP маркера. Осим тога, због изражене селективности, истовременом применом комбинације више специфичних прајмера у оквиру једне анализе може се квантификовати полиморфизам већег броја локуса (Karp и сар., 1996; Kadirvel и сар., 2015).

Микросателити представљају кратке (обично дужине један до шест нуклеотида), некодирајуће, тандемски поновљене низове нуклеотида који су присутни у променљивом броју копија у локусу и широм генома еукариота. Такође се називају и поновци једноставних секвенци (*Simple Sequence Repeats*; SSR), кратки тандемски поновци (*Short Tandem Repeats*; STR), полиморфизам дужине једноставних секвенци (*Simple Sequence Length Polymorphisms*; SSLP), или микросателитска места означена секвенцом (*Sequence-Tagged Microsatellite Sites*; STMS). Ако су репетитивни поновци дужине од десет до неколико стотина нуклеотида називају се минисателити или варијабилни број тандемских поновака (*Variable Number of Tandem Repeats*; VNTR). Микросателити и минисателити су мултилокусни, кодоминантни маркери које одликује висока стопа мутације. Веома су варијабилни, често и са десетинама алела у једном локусу. Алели се међусобно разликују у дужини која зависи од броја понављања одређене нуклеотидне секвенце. Полиморфизам ових локуса се детектује применом PCR методе и прајмера који обележавају репетитивну секвенцу ДНК са обе стране, након чега се варијабилност у дужини дате секвенце визуелизује на гелу. Такође, квантификовање варијабилности микросателитских локуса је употребом флуоресцентно обележених прајмера и ласерске детекције секвенцима аутоматизовано, а пропустност (истовремено процесуирање већег броја узорак/реакција) значајно повећана. Последњих десетак година представљају важну методу за проучавање генетичког диверзитета, а генотипизација јединки се обично заснива на анализи више микросателитских локуса (Agarwal и сар., 2008; Hall и Beissinger, 2014; Kadirvel и сар., 2015).

Полиморфизам појединачних нуклеотида (*Single Nucleotide Polymorphism*; SNP) се детектује секвенцирањем одабраних региона у геному и односи се на варијабилност појединачних нуклеотида у геному узроковану тачкастим мутацијама. Дакле, SNP (постало је уобичајено да се у говору користи израз “снп”/”снпови”) је разлика у нуклеотиду на одређеној позицији у ДНК секвенци између јединки, при чему алтернативне нуклеотидне базе на датој локацији представљају различите алеле унутар локуса. SNP су кодоминантни маркери које одликује висок ниво полиморфизма. Осим тога, бројни су и широко распрострањени по читавом геному (са различитим фреквенцијама присутни су како у кодирајућим и некодирајућим регионима гена, тако и у интергенским регионима). Већина SNP се налази у некодирајућим регионима генома, те не утичу на фенотип (селективно неутрални маркери). Међутим, неки SNP представљају мутације у секвенци унутар гена или у регулаторном региону гена, што може довести до промене у структури и функцији протеина који дати ген кодира

(селективно адаптивни маркери) (Agarwal и сар., 2008; Hall и Beissinger, 2014; Kadirvel и сар., 2015).

Southern blotting представља тест хибридизације испитиване ДНК и обележених ДНК проба. Фрагменти испитиване ДНК (добијени сечењем ДНК помоћу рестрикционих ензима) се прво раздвоје на гелу методом електрофорезе, а потом са гела пренесу на специјално дизајнирану нитроцелулозну мембрану и инкубирају са обележеним пробама. Уколико секвенца ДНК пробе препозна себи комплементаран низ нуклеотида фрагмента испитиване ДНК, доћи ће до њихове хибридизације, а настали хибриди ће бити визуелизовани на гелу у виду трака.

ДНК пробе су делови једноланчане ДНК који се користе за детектовање присуства комплементарних секвенци у испитиваној ДНК техником хибридизације. ДНК пробе могу бити радиоактивно или нерадиоактивно обележене тако да се фрагменти ДНК који показују хомологију детектују коришћењем ауторадиографије.

Ланчана реакција полимеразе (*Polymerase Chain Reaction*; PCR) је хемијски процес у ком се формира велики број копија одређеног региона или локуса испитиване ДНК помоћу кратких једноланчаних делова ДНК који се називају **прајмери** (*primer*). Дакле, у *in vitro* условима, у једној микротуби која садржи смешу свих неопходних елемената (ДНК, пар прајмера који дефинишу регион ДНК ланца који ће се амплификовати, ензим ДНК полимеразе, слободни нуклеотиди, одговарајући пуфер са магнезијум-хлоридом) врши се репликација жељене секвенце ДНК молекула. PCR се састоји од низа циклуса (обично између 20 и 35); у току сваког циклуса се одвија денатурација дволанчане ДНК и формирање једноланчаних ДНК молекула, потом хибридизација прајмера и једноланчане ДНК, те елонгација прајмера и полимеризација ДНК ланца комплементарног секвенци региона једноланчане ДНК обележеног прајмерима. Дакле, прајмери означавају места од којих започиње синтеза новог ДНК ланца на основу матрице једноланчане ДНК. PCR реакција се одвија у уређају (PCR машина) који обезбеђује неопходне температурне услове за одвијање сваке етапе циклуса.

Секвенцирање је техника којом се идентификују нуклеотидне базе и њихов редослед дуж ланца ДНК или РНК. До сада су развијене различите врсте технологија секвенцирања (за више детаља видети у Heather и Chain, 2016; Gasparyan и сар., 2019). Аутоматизацијом секвенцирања процес је знатно убрзан уз смањење трошкова. Аутоматско секвенцирање се заснива на техници обележавања нуклеотида флуоресцентним бојама које се стимулишу помоћу ласера, а флуоресцентна емисија се детектује на уређају са наелектрисањем који је у стању да одреди таласну дужину. Дакле, читавају се различити интезитети и таласне дужине боје које одговарају различитим нуклеотидима.

Табела 13. Преглед основних карактеристика најчешће коришћених маркера у анализи генетичког диверзитета. Извор: модификовано из Hall и Beissinger (2014), Nadeem и сар. (2018).

	Алозими	RFLP	RAPD	AFLP	Микросателити	SNP
Метода детекције	електрофореза	хибридизација	PCR	PCR	PCR	секвенцирање
Заступљеност у геному ¹	мала	висока	висока	висока	средња/висока	средња/висока
Ниво полиморфизма ²	низак	средњи	средњи	средњи	висок	висок
Ефекат селекције	неутралан	неутралан	неутралан	неутралан/адаптивни	неутралан	неутралан/адаптивни
Однос алела	кодоминантан	кодоминантан	доминантан	кодоминантан/доминантан	кодоминантан	кодоминантан
Поновљивост	висока	висока	ниска	средња/висока	висока	висока
Ниво техничке захтевности	низак	висок	низак	средњи	низак/средњи	низак
Количина потребне ДНК	-	велика	мала	средња	мала	мала
Аутоматизација ³	не	Не	да	да	да	да

¹Број маркера који се могу генерисати је одређен учесталошћу којом се места од интереса јављају у геному. У погледу покривености генома, неки маркери су прилично равномерно распоређени по геному, док други могу бити груписани у одређеним регионима; ²Моћ резолуције маркера је одређена нивоом детектованог полиморфизма, а који је у корелацији са стопом мутације на датим местима у геному; ³Технике које се могу аутоматизовати су веома пожељне због потенцијала за велику пропусност узорака (истовремено процесуирање већег броја узорака) чиме се убрзава процес и смањују трошкови. RFLP - полиморфизам дужине рестрикционих фрагмената; RAPD - случајно умножена полиморфна ДНК; AFLP - полиморфизам дужине умножених фрагмената; SNP - полиморфизам појединачних нуклеотида; PCR - ланчана реакција полимеразе.

8.2. Типови молекуларних података

Описивање генетичке варијабилности и процена генетичке диференцијације између популација или јединки узоркованих у одређеном географском пределу подразумева прикупљање генетичких података, њихову обраду (трансформацију) у смислу израчунавања генетичких мера (= параметара), те статистичку анализу података. Облик прикупљених података ће зависити од одабраног типа ДНК и локуса. Тако, на пример, у случају мтДНК и цпДНК које се углавном анализирају применом метода RFLP или ДНК секвенцирања подаци су у форми хаплотипа, јер се ови геноми унипарентално наслеђују (стога су и њихови локуси хаплоидни). Дакле, хаплотип је хаплоидна верзија генотипа и у сету података нуклеотидних секвенци мтДНК и цпДНК се односи на јединствену ДНК секвенцу. Са друге стране, формат генетичких података за нДНК зависиће од односа алела у локусу. Наиме, кодоминантни маркери (попут микросателита и SNP) омогућавају одређивање генотипова, хомозиготних и хетерозиготних, и учесталости алела у локусима, док се код доминантних маркера (попут AFLP) присуство оба алела у одређеном локусу (хетерозигот) не може идентификовати. Образац трака на гелу добијен за доминантне маркере/локусе се интерпретира у бинарном формату, односно бележи се присуство (означава се са 1) или одсуство (означава се са 0) трака, тј. фрагмената одређене величине и хетерозиготност не може директно утврдити.

9. Мерење генетичке варијабилности

Прикупљени генетички подаци се могу користити необрађени (на пример, у формату генотипова), или у облику неке од генетичких мера, а које могу бити засноване на јединци или популацији. Одабир нивоа анализе генетичких података (индивидуални или популациони) зависиће од просторне дистрибуције анализираног таксона. Популациони приступ се обично примењује за врсте код којих су јединке дистрибуиране у дискретним групама, а индивидуални приступ када је врста континуирано дистрибуирана у пределу и не постоје јасне границе између популација (Hall и Beissinger, 2014; Waits и Storfer, 2016).

Разликујемо више типова генетичких мера: дескриптивне, мере протока гена (миграције) и дистанце. Дескриптивне мере описују основну генетичку варијабилност популације, мерама миграције се израчунава стопа миграције или број миграната који се остварује између популација указујући колика је повезаност анализираних популација, док се мерама дистанце квантификује генетичка диференцијација јединки и популација (Hall и Beissinger, 2014). Многе од генетичких мера се могу израчунати употребом различитих маркера, мада постоје и параметри који су погодни само за

одређене маркере. Значи, одабир аналитичке методе зависи и од врсте коришћеног генетичког маркера.

Када је у предеоној генетици јединица анализе генетичких података популација, из групе дескриптивних генетичких мера за квантификовање генетичке варијабилности на основу кододоминантних локуса нДНК (алозими, микросателити, SNP) најчешће се користе следећи параметри:

- фреквенција генотипова и алтернативних алела у локусу;
- хетерозиготност (H) - израчунавају се уочена и очекивана хетерозиготност. Уочена хетерозиготност (H_o) представља удео хетерозиготних јединки у анализираном узорку. Очекивана хетерозиготност (H_e) је процена хетерозиготности (удео хетерозиготних јединки у популацији) која се очекује у популацији према Харди-Вајнберговом закону, а на основу фреквенција регистрованих алела. Често се назива и диверзитет гена. За обе мере израчунавају се вредности за сваки локус појединачно, а потом и просечне вредности по локусу;
- полиморфност (P) - представља удео полиморфних локуса у укупном броју анализираних локуса. Локус се може дефинисати као полиморфан ако фреквенција најчешћег алела није већа од 99% (или 95% према другом критеријуму);
- диверзитет алела (A) - представља просечан број алела по локусу у популацији. Мера је веома осетљива на разлике у величини узорака, па се често замењује мером алелно богатство која укључује корекцију за разлике у величини популација;
- алелно богатство (A_r) - просечан број алела по локусу у популацији независно од величине популације. Ова мера је најосетљивија за откривање недавних губитака генетичког диверзитета. Може се израчунати и користећи најмање бројну популацију као референтну (*rarefaction*);
- присуство јединствених/приватних алела (*unique, private, u*; алели присутни само у одређеној популацији једне врсте) и ретких алела (*rare, r*; алели чија је фреквенција мања од 0,10).

За доминантне локусе (RAPD, AFLP) као стандардна мера диверзитета користи се просечан број трака по популацији, док се мере аналогне очекиваној хетерозиготности (H_e) и диверзитету нуклеотида (π) такође могу израчунати под одређеним претпоставкама (Waits и Storfer, 2016).

За генетичке податке у облику нуклеотидних секвенци мтДНК и цпДНК основне мере диверзитета су:

- број хаплотипова - укључује апсолутни број хаплотипова, као и број јединствених/приватних хаплотипова;

- диверзитет хаплотипова (h) - вероватноћа да су два насумично узоркована хаплотипа различита. Назива се и диверзитет гена и представља хаплоидну верзију очекиване хетерозиготности;
- диверзитет нуклеотида (π) - просечан број нуклеотидних разлика по месту између било која два случајно одабрана хаплотипа;

Генетички диверзитет унутар самог генотипа (тј. на нивоу индивидуе) се квантификује као индивидуална хетерозиготност (H) која представља удео хетерозиготних локуса у укупном броју анализираних локуса јединке. Овај параметар је осетљив на разлике између локуса у броју или фреквенцији алела или када није извршена генотипизација за исти скуп локуса за све јединке у узорку. Зато су развијене и друге, алтернативне мере: стандардизована хетерозиготност (SH ; хетерозиготност сваког локуса пондерисана са просечном хетерозиготношћу датог локуса), унутрашње сродство (*internal relatedness*, IR ; израчунава пропорцију хомозиготних локуса пондерисано на основу фреквенције алела) и хомозиготност по локусу (HL ; индекс хомозиготности који пондерише допринос локуса са њиховом очекиваном хетерозиготношћу) (Waits и Storfer, 2016).

10. Процена генетичке структуре популације и детектовање баријера

Појам генетичка структура популације (= популациона структура, подела популације) односи се на обрасце дистрибуције адаптивно неутралне генетичке варијабилности у простору и времену, а који настају када у популацији дође до одступања од панмиксије (било да се то десило у прошлости или сада). Наиме, у популацији у којој влада случајно укрштање (= панмиксија; свака јединка у популацији има једнаку вероватноћу да ступи у репродуктивне односе са било којом јединком супротног пола дате популације) очекује се да фреквенције алела буду приближно једнаке између локалних популација (= субпопулација, дема), а проток гена неограничен. Проток гена представља трансфер генетичког материјала из једне популације у другу, а постиже се кретањем (= миграција) јединки и њиховом репродукцијом. Међутим, географске одлике предела и средински фактори могу условити да између локалних популација буде смањен проток гена, што временом доводи до образовања специфичних образаца дистрибуције генетичке варијабилности унутар и између популација, односно популационе структурираности. Као пример, можемо замислити једну врсту чије јединке насељавају више острва која су међусобно мање или више удаљена. Размена јединки (миграција) између острвских популација може бити ограничена (одређени број јединки мигрира између острва с времена на време), или потпуно прекинута присуством физичке баријере (море, удаљеност острва), па се еволуционе промене у свакој острвској популацији одвијају у одређеном степену независно (или потпуно независно у сличају одсуства протока гена). То резултира

генетичком диференцијацијом између њих. Дакле, генетичку структуру популације карактерише одређени број субпопулација унутар ње, фреквенције генетичких варијанти (алела) у свакој субпопулацији (генетичка варијабилност унутар локалних популација), као и степен генетичке изолованости субпопулација (генетички диверзитет између локалних популација).

Развијено је више метода за утврђивање генетичке структуре и квантификовање дивергенције између популација. Уопштено говорећи, ове методе се могу поделити у две групе засноване на концептуално различитим приступима у анализи. Први, традиционални, приступ у проучавању генетичке структуре популације заснива се на утврђивању популационих генетичких разлика, при чему су популације *a priori* (унапред) дефинисане. То значи да се на самом почетку анализе узорковане јединке распоређују у групе (популације) утврђене према одређеном критеријуму (на пример, месту узорковања; све јединке сакупљене на једном локалитету чине популацију), а потом тестира статистичка значајност разлика у генетичкој конституцији између популација. Међутим, ако јасно дефинисане и дискретне *a priori* популације не постоје или немамо разлога да очекујемо да су оне реалне (јер су, на пример, јединке континуирано дистрибуиране у пределу), користи се други приступ који не захтева да припадност јединки групама буде унапред одређена. Овај приступ утврђивања генетичке структуре подразумева идентификовање панмиктичних јединица у укупном узорку. Наиме, након генотипизације јединки, у узорку се на основу индивидуалних генетичких разлика издвајају групе генетички сличних јединки (генетички кластери).

За евалуацију генетичке структуре и протока гена у природним популацијама се обично користе: параметри генетичке структуре, методе кластер анализе и методе за детекцију баријера.

10.1. Параметри генетичке структуре популације

Параметри генетичке структуре обухватају низ мера које указују на степен повезаности популација (односно, проток гена између њих; мере миграције) и ниво генетичке диференцијације између популација или јединки исте врсте (мере дистанце). Наиме, што је ниво диференцијације нижи, значи да је генетичка сличност између популација већа. Такође, када је током дужег временског периода између изолованих популација размена јединки мала (или је уопште нема) повећава се ниво генетичке диференцијације.

Процена генетичке структуре и степена протока гена између популација се обично заснива на следећим генетичким мерама:

- F_{ST} - најчешће коришћена мера диференцијације између популација; један је од три параметра Рајтове F статистике (F_{IT} , F_{IS} , F_{ST} ; Wright, 1951). F статистика описује поделу генетичке варијабилности унутар врсте, односно између тоталне

популације (обухвата све популације врсте које се разматрају; T), субпопулације (S) или јединки унутар субпопулације (I). Како структурираност популација и инбридинг узрокују на редукцију хетерозиготности, параметри F статистике се могу квантификовати на основу хетерозиготности израчунате на различитим нивоима популационе структуре (јединка, субпопулација, тоталне популација). Тако, F_{ST} параметар представља смањење просечне хетерозиготности субпопулација (H_S) у односу на очекивану хетерозиготност тоталне популације (H_T):

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

Дакле, F_{ST} параметар квантификује степен међупопулационе диференцијације и може имати вредност од 0 (панмиксија; две популације имају идентичне фреквенције алела) до 1 (потпуна генетичка изолација субпопулација, две популације не деле заједничке гене). Примењује се за описивање генетичке структуре популације на основу диплоидних генетичких података (алозими, нДНК секвенце), али се може модификовати (уместо хетерозиготности користи се диверзитет хаплотипова) за примену на хапloidним подацима (мтДНК, цпДНК). За процену генетичке диференцијације у употреби су и други параметри, попут G_{ST} , R_{ST} и тета (θ) параметара;

- G_{ST} - представља проширење F_{ST} параметра; примењује се када се у локусу налази више од два алела и процењује удео диверзитета гена који је распоређен између популација у односу на укупни диверзитет гена;
- R_{ST} - мера аналогна F_{ST} параметру, прилагођена је за анализу микросателитних података са великим бројем алела и хетерозиготношћу. Узима у обзир разлику у броју поновака између алела;
- тета (θ) - мера аналогна F_{ST} параметру заснована на варијансама фреквенције алела. Може се израчунати на основу хомозиготности (θH), броја сегрегационих (полиморфних) места (θS), броја алела (θk) и просечаног броја нуклеотидних разлика ($\theta \pi$). Ова мера узима у обзир разлике у величини између популација, а не захтева претпоставке о броју узоркованих популација, величини узорка или хетерозиготности локуса;
- анализа молекуларне варијансе (*Analysis of molecular variance*; AMOVA) - представља непараметарску статистичку методу аналогну традиционалној анализи варијансе (ANOVA). Омогућава хијерархијску поделу укупне генетичке варијабилности на компоненте различитих нивоа организације узорка (на пример, унутар популација, између популација унутар региона, између различитих региона) и процену параметара F статистике (или њених аналога) који одражавају диверзитет на различитим нивоима поделе. AMOVA се може применити на било који облик молекуларних података (укључујући кододоминантне, бинарне, хапloidне маркере и ДНК секвенце), а сума квадрата одступања (SS) и средина квадрата одступања (варијанса; MS) се израчунавају на основу матрице генетичке дистанце. Из односа MS различитих хијерархијских

нивоа и укупне варијансе се добија одговарајући параметар F статистике (на пример, $F_{ST} = MS_{između\ populacija} / MS_{ukupno}$).

- генетичка дистанца - у употреби је више коефицијената за процену међупопулационе генетичке диференцијације; заснивају се на процени хетерозиготности или разлике у сличности алела између популација (Hall и Weissingер, 2014; Waits и Storfer, 2016);
- стопа миграције - удео имиграната у популацији након миграције;
- број миграната (Nm) - представља апсолутни број миграната који две популације размене по генерацији. Једноставан однос између F_{ST} параметра и броја миграната $F_{ST} = 1 / (1 + 4 Nm)$, где је N ефективна величина популације, а m стопа миграције се веома често користи за процену броја миграната. Алтернативни начин за процену Nm је на основу фреквенције приватних алела (Slatkin, 1985).

Генетички дисконтинуитет природних популација се на основу индивидуалних генотипова квантификује мерама дистанце:

- сродство (r ; *relatedness*) - мера идентичности (због заједничког претка) између јединки; представља генетичку сличност између две јединке у односу на ону између насумично одабраног пара јединки из неке референтне популације;
- пропорција заједничких алела (D_{PS}) - мери степен сличности у алелној идентичности између јединки;
- Брау-Суртис мера различитости (B_C) - мера позната и као разлика фреквенције алела, представља просечну апсолутну разлику у пропорцији алела између две јединке;
- a_r (Rousset, 2000) - мера заснована на изолацији дистанцом у континуирано дистрибуираним врстама.

Пример 1.

Генетичка структура популација врсте *Cheilosia urbana* (Diptera, Syrphidae) је проучена методом полиакриламид гел електрофорезе десет алозимских локуса (*Aat*, *Fum*, *Gpd-2*, *Gpi*, *Hk-2*, *Hk-3*, *Mdh-1*, *Mdh-2*, *Pgm*, *Sod*). Анализиране јединке су сакупљене на подручју Фрушке горе (Србија; CUFГ), Дубашнице (Србија; CUDUB) и Дурмитора (Црна Гора; CUDUR). На основу утврђених генотипова и њихових фреквенција, израчунате су фреквенције алела, те параметри генетичке варијабилности и диференцијације анализираних популација (Табела 14; Ludoški, 2002). Највећи броја алела (21) детектован је у популацији CUDUB, нешто мањи број у CUFГ (18), а најмањи у CUDUR (16), при чему је од укупног броја анализираних локуса шест било полиморфно. На основу просечног броја алела по локусу (A), уочене хетерозиготности (H_o) и фреквенције полиморфних локуса (P) уочава се да је популација CUDUR најмање варијабилна, док је најваријабилнија популација CUDUB.

Табела 14. Фреквенције генотипова варијабилних локуса и параметри генетичке структуре три популације врсте *Cheilosia urbana* (Diptera, Syrphidae). Значајност интерпопулационих разлика фреквенција алела тестирана је χ^2 тестом. Извор: модификовано из Ludoški (2002)

Локус	Генотип	CUFG	CUDUB	CUDUR	χ^2 тест
<i>Aat</i>	<i>b/b</i>	0,324	0,516	0,235	**
	<i>c/c</i>	0,676	0,484	0,765	
<i>Fum</i>	<i>a/a</i>	0,882	0,364	0,656	***
	<i>b/b</i>	0,118	0,636	0,344	
<i>Gpi</i>	<i>e/e</i>	0,194	0,267	0,091	ns
	<i>f/f</i>	0,694	0,500	0,818	
	<i>b/f</i>	0,028	0,067	0,061	
	<i>c/f</i>	0,083	0,167	0,030	
<i>Hk-2</i>	<i>a/a</i>	-	0,088	-	***
	<i>b/b</i>	0,973	0,824	1,000	
	<i>c/c</i>	0,027	0,088	-	
<i>Hk-3</i>	<i>a/a</i>	-	0,088	-	***
	<i>b/b</i>	0,973	0,824	1,000	
	<i>c/c</i>	0,027	0,088	-	
<i>Pgm</i>	<i>f/f</i>	-	0,185	0,032	***
	<i>g/g</i>	0,939	0,667	0,968	
	<i>h/h</i>	0,061	0,148	-	
<i>n (SE)</i>		25,7 (3,9)	23,0 (3,6)	24,6 (3,5)	
<i>A (SE)</i>		1,8 (0,3)	2,1 (0,3)	1,6 (0,3)	
<i>u</i>		-	2	-	
<i>r</i>		4	4	4	
<i>H_o (SE)</i>		0,011 (0,011)	0,023 (0,023)	0,009 (0,009)	
<i>H_e (SE)</i>		0,126 (0,053)	0,261 (0,075)	0,112 (0,055)	
<i>P_(0,95)</i>		0,400	0,600	0,300	
<i>P_(0,99)</i>		0,600	0,600	0,400	

n - просечна величина узорка по локусу; *A* - просечан број алела по локусу; *u* - број јединствених алела; *r* - број ретких алела; *P_(0,95)* - фреквенција полиморфних локуса на основу критеријума 0,95; *P_(0,99)* - фреквенција полиморфних локуса на основу критеријума 0,99; *H_o* - просечна уочена хетерозиготност; *H_e* - просечна очекивана хетерозиготност; *SE* - стандардна грешка; *** *P* < 0,001; ** *P* < 0,01; ns - није статистички значајно

Такође, присуство јединствених (приватних; *u*) алела регистровано је једино у популацији CUDUB, док су ретки алели присутни у све три анализирани популације (Ludoški, 2002). Анализом међупопулационе генетичке диференцијације нису регистровани локуси који указују на потпуну генетичку различитост популација, док је тестирањем значајности разлика фреквенција алела шест варијабилних локуса утврђена статистички значајна разлика између популација ($\chi^2 = 135,23$; $P < 0,001$). Такође, вредност F_{ST} параметра указује да је степен генетичке диференцијације између популација умерен ($F_{ST} = 0,115$). Израчуната вредност генетичке дистанце (D ; Nei, 1972) између парова популација креће се у опсегу $D = 0,009 - 0,047$ (Ludoški, 2002).

Пример 2.

На основу нуклеотидне секвенце гена за цитохром *b* митохондријског генома (*cytb* мтДНК) јединки узоркованих на четири локалитета Дабарског поља (Босна и Херцеговина) проучена је популациона структура рибе *Telestes dabar* (Leuciscidae) (Francuski и сар., 2019). Анализом 47 секвенци идентификовано је укупно пет хаплотипова, од којих су три била заједничка за све четири популације, а по један приватни (јединствени) хаплотип је регистрован у две популације (Табела 15; Francuski и сар., 2019). Забележене су високе вредности генетичког диверзитета (h) и ниске вредности нуклеотидног диверзитета (π) (Табела 15). Анализа молекуларне варијансе (AMOVA) и F_{ST} вредност указују на одсуство генетичке диференцијације између анализираних популација ($F_{ST} = 0,008$; $P > 0,05$); наиме, готово целокупна варијабилност (99,14%) дистрибуирана је унутар популација (Табела 16; Francuski и сар., 2019).

Табела 15. Параметри генетичког диверзитета популација врсте *Telestes dabar* (Leuciscidae) на основу секвенци *cytb* мтДНК. Извор: модификовано из Francuski и сар. (2019)

Локалитет	N	S	H	H_P	h	π	k
Опачица	14	5	3	-	0,692	0,00190	2,077
Прибитул	15	6	4	1	0,714	0,00204	2,229
Сушки поток	9	5	4	1	0,750	0,00157	1,722
Вријека	9	5	3	-	0,667	0,00223	2,444
Укупан узорак	47	6	5	-	0,702	0,00198	2,163

N - број јединки/секвенци, S - број варијабилних места, H - број хаплотипова, H_P - број приватних хаплотипова, h - диверзитет хаплотипова, π - диверзитет нуклеотида, k - просечан број нуклеотидних разлика

Табле 16. Анализа молекуларне варијансе (AMOVA) на основу *cytb* мтДНК секвенци популација врсте *Telestes dabar* (Leuciscidae). Извор: модификовано из Francuski и сар. (2019)

Извор варијације	Степени слободe	Сума квадрата	Компоненте варијабилности	Процент варијабилности	F_{ST}
Између популација	3	2,91	0,007	0,86	0,008
Унутар популација	43	37,89	0,881	99,14	
Укупно	46	40,80	0,888		

Последњих двадесетак година развијен је низ програмских пакета намењених извођењу основних анализа популационе генетике, попут Arlequin (Excoffier и Lischer, 2010), DnaSP (Librado и Rosas, 2009), FSTAT (Goudet, 2002), GENEPOP (Rousset, 2008), GENETIX (Belkhir, 2004), више R пакета - *gstudio* (Dyer, 2014), *pegas* (Paradis и сар., 2015), *genetics* (Warners, 2015) и други. Међу њима, најчешће коришћени су:

- Arlequin (Excoffier и Lischer, 2010) интегрише више основних и напредних метода за анализу неколико типова генетичких података, било да су у облику хаплотипова или генотипова (ДНК секвенце, RFLP и микросателитски подаци, фреквенције алела), или матрице израчунатих генетичких дистанци. Програмом су обухваћене методе за анализу образаца генетичког диверзитета унутар популација (попут израчунавања параметара генетичког диверзитета, фреквенција алела и хаплотипова, параметара демографске или просторне експанзије, гаметске неравнотеже, одступања од Харди-Вајнбергове равнотеже, тестови неутралности) и између популација (детектовање заједничких хаплотипова, AMOVA, израчунавање параметара F статистике, генетичке дистанце између популација, тестови диференцијације популација, Мантелов тест корелације матрица дистанце).
- DnaSP (Librado и Rosas, 2009) је програм за анализу полиморфизма ДНК на основу нуклеотидних секвенци. Програм израчунава неколико мера варијабилности ДНК секвенце унутар и између популација на некодирајућим, синонимним или несинонимним местима. Такође, израчунава степен гаметске неравнотеже, рекомбинације, проток гена, параметре конверзије гена, а обухвата и више тестова неутралности.

10.2. Методе за идентификовање просторних образаца дистрибуције генетичког диверзитета

Најчешће коришћени приступи за квантификовање генетичке структуре популација су методе генетичког додељивања (*genetic assignment methods*) и методе груписања (*clustering methods*). Методе додељивања обухватају неколико сродних статистичких метода које користе генетичке информације како би се утврдила припадност јединки (или група јединки) некој популацији. Једна од њих је тест додељивања (*assignment test*), широко заступљени статистички тест хипотезе да мултилокусни генотип анализирани јединке потиче из одређене популације (Manel и сар., 2005). Дакле, тест додељивања настоји да распореди (додели) јединке у њихове популације порекла, а на основу генотипа и очекиваних вероватноћа да ће се тај генотип појавити у свакој од потенцијалних изворних популација. Очекиване генотипске вероватноће се израчунавају на основу узорка из сваке потенцијалне изворне популације, а генотип се потом додељује оној популацији која има највећу вероватноћу да ће се дати генотип појавити у њој (Manel и сар., 2005). Овакав приступ класификовања јединки у групе заснован на вероватноћи подразумева да су потенцијалне изворне популације унапред дефинисане, случајно узорковане, те да се налазе у Харди-Вајнберговој и гаметској равнотежи (Manel и сар., 2005; Waits и Storfer, 2016). Поред традиционалног теста додељивања, у употреби су и анализа родитељства (*parentage analysis*; метода класификације којом се на основу мултилокусних генотипова идентификују родитељи јединке или групе јединки, и потом се те јединке групишу заједно са родитељима) и анализа генетичког мешања (*genetic mixture analysis*; процењује колики удео у мултилокусном генотипу сваке јединке је допринос различитих изворних популација, а додељивање одређеној популацији је базирано на вероватноћи) (Manel и сар., 2005). Тако се детектовањем миграната (или њихових потомака), као и родитеља потомака, на директан начин може документовати и кретање јединки и проток гена између популација (Waits и Storfer, 2016). На пример, удео јединки једне популације које су додељене у другу популацију тумачи се као процена протока гена из друге у прву популацију.

У природи, јединке одређене врсте су обично груписане у локалне популације. Некада је структурираност популације лако претпоставити на основу географске локације узоркованих јединки. Наиме, тестирањем нулте хипотезе да су јединке узорковане на различитим локалитетима део једне панмиктичне популације закључује се да ли постоји популациона структурираност. Када тест додељивања покаже да јединке имају већу вероватноћу да буду додељене оној групи која одговара месту њиховог узорковања него било којој другој, може се закључити да постоји популациона структурираност. Међутим, границе између географски дефинисаних популација нису увек баш јасне, што отежава, али и доводи у питање оправданост употребе традиционалног приступа *a priori* дефинисања популација. Заправо, дефинисање популација унапред може бити прилично субјективно јер је засновано, на пример, на физичким карактеристикама или месту узорковања јединки и тешко је знати да ли

груписање јединки према овим критеријумима одражава формирање група и у генетичком смислу (Pritchard и сар., 2000). Стога, за процену генетичке структуре популација користе се методе кластер анализе. Ове методе груписања базирају се на препознавању група (кластера) генетички сличних јединки унутар велике популације проценом њиховог генетичког порекла (François и Waits, 2016). Основна претпоставка алгоритама за груписање је да постоје биолошки значајни кластери које треба открити.

Статистичке методе за утврђивање структуре популације и процену генетичког порекла без претходног дефинисања популација обухватају две основне групе метода (Jombart и сар., 2010; Tonkin-Hill и сар., 2019):

1) методе засноване на дистанци (*distance-based methods*) - користе експлораторни приступ анализи података (*exploratory data analysis*); другим речима, ове методе сумирају скупове података мутилокусних генотипова помоћу мањег броја варијабли које потом користе за издвајање генетичких кластера без претпоставки о процесима који генеришу податке. У ту сврху користе се стандардне мултиваријантне статистичке методе (на пример, факторска анализа, коресподентна анализа, технике мултидимензионалног скалирања, регресија и друге) од којих је анализа главних компоненти (PCA) најчешће заступљена;

2) методе засноване на моделима (*model-based methods*) - користе моделе популационе генетике који описују дистрибуцију фреквенција алела у структурираним популацијама. Методе препознају генетичке кластере на основу мутилокусних генотипова и потом додељују јединке идентификованим кластерима. Израчунава се вероватноћа да јединка припада одређеном кластеру, а модели могу дати и процену непоузданости за утврђене параметре груписања. Статистичко закључивање се у моделима популационе структуре (односно, закључивање о параметрима сваког кластера, као и о припадности јединки групама) обично заснива на Бајесовом приступу или на максималној вероватноћи (Pritchard и сар., 2000; Tonkin-Hill и сар., 2019). Такође, многи модели груписања утврђују и колики је допринос популација предака у посматраном узорку исказујући га коефицијентима порекла јединки (*ancestry coefficients*). Претпоставка је да су алели наслеђени из две или више родитељских популација, а да су родитељске популације остаци предачких популација (François и Waits, 2016).

Дакле, методама обе групе могу се израчунати индивидуални коефицијенти порекла или предвидети припадност јединки одређеним генетичким кластерима, а генетичка структура популације визуелизовати графичким приказима.

10.2.1. Анализа главних компоненти

Просторни обрасци генетичке варијабилности могу се идентификовати применом метода мултиваријантне статистике. Анализа главних компоненти (PCA) је стандарна статистичка метода која је своју примену пронашла и у популационој генетици. Анализа сумира скупове мултилокусних генотипских података редукујући димензионалност података и омогућава сагледавање дистрибуције варијабилности у простору². Графичко приказивање резултата PCA олакшава и њихово интерпретирање. Стандарно, резултати се приказују у виду графика расејања (*scatterplot*) који приказује простор дефинисан главним компонентама (обично, са прве две главне компоненте - PC1 и PC2, мада и нека наредна може бити такође информативна) у који су пројектовани појединачни генотипови. Сагледавањем дистрибуције јединки и удаљености између њих тумачи се популациона структура анализираних узорака. На пример, на генетичку структурираност популације указује груписање јединки/генотипова у дискретне групе (генетичке кластере), док преклапање кластера може бити резултат заједничког порекла. Узимајући у обзир географске локације анализираних јединки генетичка варијабилност се може у простору приказати и у виду синтетичке мапе. Ове мапе настају интерполацијом скорова главних компоненти и приказивањем тих вредности на географској мапи (Manel и сар., 2003; François и Waits, 2016).

Просторна анализа главних компоненти (*spatial PCA*; sPCA) представља модификацију стандардног PCA алгорита и намењена је за анализу просторно експлицитних (геореференцираних) генетичких података. Просторна PCA генерише нови скуп варијабли (главних компоненти) које, за разлику од стандардне PCA, максимизирају и генетичку варијансу и просторну аутокорељацију квантификовану Морановим *I* параметром (Jombart и сар., 2008; François и Waits, 2016; о просторној аутокорељацији видети касније). Такође, на основу алелних фреквенција јединки или популација, sPCA сумира генетичку варијабилност и просторну структурираност јединки (или популација), те омогућава разликовање два типа образаца структуре: глобални (када се разликују просторно одвојене групе јединки/острва, клине; показује позитивну просторну аутокорељацију) и локални (изражене су генетичке разлике између суседа; показује негативну просторну аутокорељацију) (Jombart и сар., 2008). Визуелизацијом индивидуалних PC скорова у виду графика расејања омогућава детектовање присуства генетичког дисконтинуитета.

² За више информација о методи видети у делу 5.3. Мултиваријантне статистичке методе

10.2.2. Методе Бајесове кластер анализе

Најчешће коришћена метода заснована на моделу је Бајесова кластер анализа. Бајесов алгоритам је имплементиран у многе компјутерске програме за популациону генетику, а међу којима се својом широком употребом посебно издвајају Structure, BAPS, Geneland, Tess. Уопштено говорећи, ови програми имају за циљ да открију имигранте у узорцима за које су анализирани различити мултиалелни маркери, а базирајући се на чињеници да ће имигранти имати различите мултилокусне генотипове у односу на оне очекиване за нативне јединке (Excoffier и Heckel, 2006). Методе Бајесове кластер анализе су изведене из традиционалних тестова додељивања; наиме, на основу мултилокусних генотипова, јединке или делови њихових генома се додељују низу кластера, при чему групе (кластери) нису унапред дефинисане, већ се морају утврдити на основу датих података. Дакле, Бајесове методе груписања настоје да у скупу анализираних јединки креирају групе, панмиктичне скупове јединки (популације) у којима су Харди-Вајнбергова неравнотежа и гаметска неравнотежа минимизирани (одступање од Харди-Вајнбергове равнотеже и гаметска неравнотежа води до поделе популације на субпопулације). Потом, додељују сваки појединачни генотип у потенцијалне популације његовог порекла (претпостављене кластере), док истовремено процењују постериорне дистрибуције за параметре модела. Претпоставка ових метода је да посматрани подаци потичу из модела вероватноће са непознатим параметрима, а њихов циљ је да користећи постојеће податке процене параметаре модела и степен непоузданости ових процена. Значи, поређењем постериорних вероватноћа за различите бројеве претпостављених кластера (у пракси, у компјутерским програмима за Бајесову кластер анализу истраживач дефинише највећи број претпостављених кластера) закључује се о броју генетичких кластера (популација) у анализираном узорку. Овај сложени прорачун постериорних вероватноћа се може извести помоћу Марковљев ланац Монте Карло алгоритма (*Markov Chain Monte Carlo*; МСМС) (Pritchard и сар., 2000; Manel и сар., 2005).

Бајесова анализа је метода статистичке анализе која полази од претходних дистрибуција за параметре модела, ажурира их на основу посматраних података и долази до постериорне дистрибуције вероватноће. Подела анализираног узорка на групе (кластере) постаје параметар модела вероватноће за дате податке, подложна је претходним претпоставкама, а закључак о груписању се изводи из својстава постериорне дистрибуције за дате параметре модела.

Марковљев ланац Монте Карло алгоритам (*Markov Chain Monte Carlo*; МСМС) је техника симулације за генерисање узорака из дистрибуције вероватноће од интереса. Ова техника се често користи за процену заједничке постериорне дистрибуције скупа параметара без потребе да се истражује цео параметарски простор (скуп свих допустивих вредности параметра). Наиме, уместо да исцрпно истражује цео параметарски простор, МСМС истражује онај део простора са великом вероватноћом.

МСМС процењује вероватноћу за различите тачке (стање) у датом простору са циљем да ефикасно истражити параметарски простор и добије узорак вредности параметара који ће бити добар репрезент праве постериорне дистрибуције (Excoffier и Heckel, 2006). МСМС може проценити сложене мултиваријантне дистрибуције које се не могу генерисати стандардним методама симулације.

Харди-Вајнбергова равнотежа (*Hardy-Weinberg equilibrium*) - полазећи од учесталости алела добијене фреквенције различитих генотипских класа се не разликују од оних очекиваних под претпоставком случајног укрштања (панмиксија) у идеалној популацији. Другим речима, релативне фреквенције алела и генотипова из генерације у генерацију остају непромењене. На одступање од Харди-Вајнбергове равнотеже може утицати селекција, миграције или структурираности популације.

Гаметска неравнотежа (*linkage disequilibrium*) - неслучајна асоцијација алела два или више локуса; неке комбинације алела различитих локуса имају тенденцију да се у популацији појављују знатно чешће него што је очекивано према Менделовом закону слободног комбиновања алела.

Примена Бајесовог алгоритма за груписање омогућава и избор једног од два генетичка модела: модел са мешањем и модел без мешања. Модел без мешања (*mixture* или *no/without admixture*) претпоставља да се анализирани узорак састоји од K дивергентних генетичких кластера (популација) и да свака јединка потиче из једне популације. Стога се јединке на основу вероватноће додељују у један од K кластера, а утврђене вероватноће се називају коефицијенти припадности (*membership coefficients*) (Pritchard и сар., 2000). Модел са мешањем (*admixture*) претпоставља да јединка носи алеле из више популација, односно, да генетички подаци потичу од мешања K претпостављених родитељских популација које можда нису доступне за истраживање. Овај модел можемо једноставно замислити као низ корпи које представљају претпостављене предачке популације (генетичке кластере) и у које се могу сместити алели неке јединке. Удео алела у свакој од тих корпи се може представити коефицијентом мешања (*admixture coefficient*) или коефицијентом порекла (*ancestry coefficients*), а који се могу израчунати за сваку јединку у узорку. Ови коефицијенти се приказују у виду матрице (Q) у којој сваки елемент (q_{ik}) представља удео генома i -те јединке пореклом из родитељске популације (k). Постериорна процена коефицијената мешања заснована је на МСМС приступу; алгоритам понавља велики број циклуса симулација фреквенције алела и параметара додељивања сваког алела, па се на основу релативног броја колико се пута алел нађе у одређеној корпи током симулације добијају постериорне процене (Corander и сар., 2008a). Дакле, у моделима са мешањем кластери се посматрају као изворне популације које су у прошлости дивергирале, достигле равнотежно стање и касније поново дошле у контакт. Овде се одређеној групи

породка не додељују јединке, већ само алели, али се коефицијенти породка могу користити за додељивање јединки у кластере (François и Waits, 2016).

У зависности да ли су у анализу укључене или не просторне информације о узоркованим јединкама (географске координате локалитета сакупљања), Бајесова кластер анализа може бити непросторна или просторна. За разлику од непросторног статистичког приступа, просторни даје увид у повезаност између генетичких и предеоних података применом тестова корелације (на пример, Мантелов тест).

Непросторна Бајесова кластер анализа (алгоритам имплементиран у програм Structure) разврстава јединке у групе (генетичке кластере) као популације (K) на основу њихових мултилокусних генотипова. Број кластера (K) одражава степен структурираности узорка. Анализа се изводи за различите вредности K , обично од $K = 1$ до $K =$ број узоркованих локалитета. Како би се проверила доследност у добијеним резултатима, за сваку вредност K се анализа понавља у више независних циклуса. Највероватнији број кластера и распоређивање јединки у кластере се може проценти на два начина: 1) одређивањем промене у граничној вероватноћи података $Pr(X/K)$ када је број кластера подешен на различите вредности; 2) употребом параметра ΔK , стопе промене функције граничне вероватноће другог реда која укључује повећање постериорних вероватноћа у распону K вредности и варијансу између независних циклуса за дати број K (Pritchard и сар., 2000; Evanno и сар., 2005). Значи, програм Structure врши додељивање јединки у кластере узастопно понављајући процес додељивања за различите бројеве кластера, а потом се помоћу софтвера Structure Harvester (Earl и vonHoldt, 2012) и CLUMPAK (Kopelman и сар., 2015) утврђује број кластера који има највећу вероватноћу и визуелизују резултати анализе.

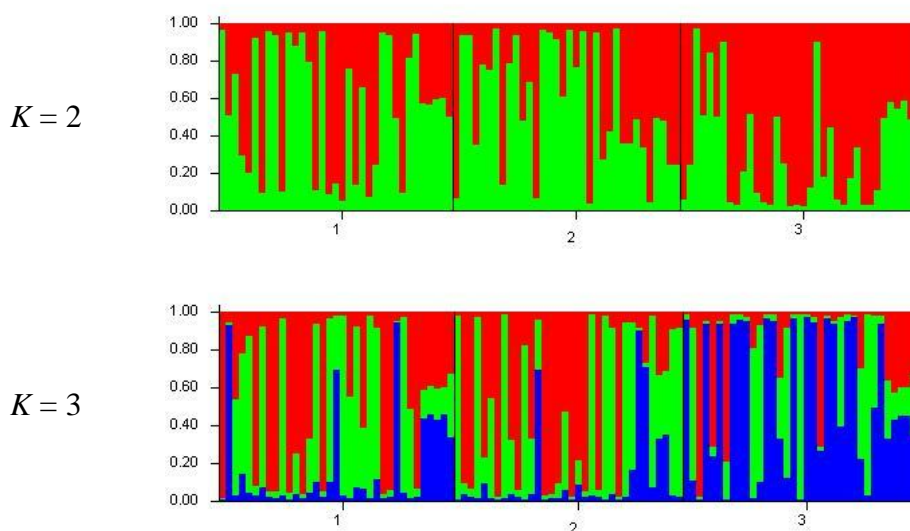
Пример 3.

Користећи непросторни приступ Francuski и сар. (2020) су у програму Structure анализирали степен популационе структурираности врсте *Eristalis arbustorum* (Diptera, Syrphidae) на основу фреквенција алела пет алозимских локуса јединки сакупљених на три локалитета у Босни и Херцеговини (Бања Лука, Приједор и Пљева). Међусобна удаљеност места узорковања износи 45 км (Бања Лука - Приједор), 59 км (Бања Лука - Пљева) и 85 км (Приједор - Пљева). На основу индивидуалних разлика јединке три анализирани популације су Бајесовом методом класификоване у два ($K = 2$) и три ($K = 3$) претпостављена кластера (Табела 17; Слика 37). Процентуална припадност јединки одређеном кластеру у оба случаја није била већа од 80% (Francuski и сар., 2020). Наиме, за $K = 2$ видимо да је већина јединки из популације Пљева (70,1%) класификована у један од два кластер (тј. у кластер I), док су јединке преостале две популације са већим процентом класификоване у други кластер (кластер II; Бања Лука 56,2%, Приједор 59,9%), што имплицира одређен степен диференцијације ове две популације у односу на популацију из Пљеве. За $K = 3$ процентуална припадност јединки анализираних популација кластерима кретала се између 17,3% и 48,7%, указујући на

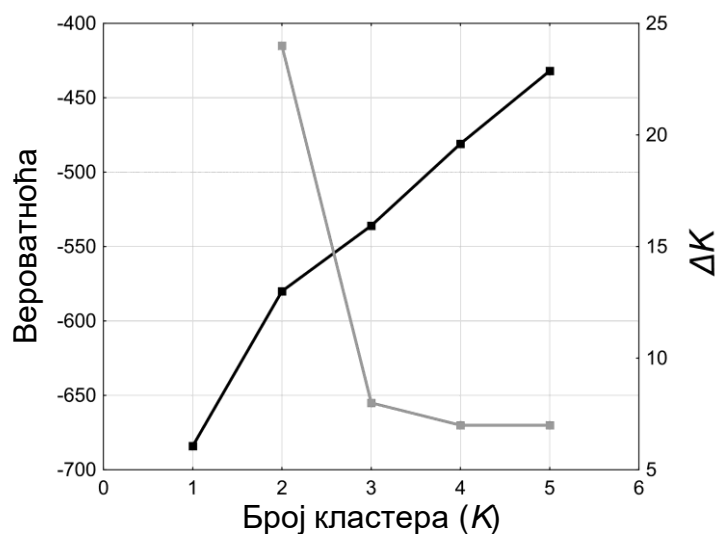
одсуство међупопулационе диференцијације (Лукач, 2014). Међутим, методом највеће вероватноће утврђено је да је највероватнији број кластера за анализирани скуп генетичких података $K = 1$ (у овом случају параметар ΔK није могао да буде примењен јер се он не може израчунати за $K = 1$) (Слика 38).

Табела 17. Процентуална припадност јединки врсте *Eristalis arbustorum* (Diptera, Syrphidae) претпостављеним кластерима за $K = 2$ (I, II) и $K = 3$ (I, II, III) утврђена *a priori* Бајесовом методом кластер анализе у програму Structure. Извор: Лукач, 2014.

	$K = 2$		$K = 3$		
	I	II	I	II	III
Бања Лука	43,8	56,2	28,7	35,6	35,7
Приједор	40,1	59,9	22,5	38,1	39,4
Пљева	70,1	29,9	48,7	17,3	34,0



Слика 37. Припадност јединки врсте *Eristalis arbustorum* (Diptera, Syrphidae) претпостављеним кластерима за $K = 2$ и $K = 3$ утврђена *a priori* Бајесовом методом кластер анализе у програму Structure. Свака вертикална линија приказује вероватноћу да је јединка да припада једном од K кластера. Различити кластери су приказани различитим бојама. 1 - Бања Лука, 2 - Приједор, 3 - Пљева. Извор: слику уступила Милица Лукач.



Слика 38. Процена вероватног броја кластера (K) у подацима врсте *Eristalis arbustorum* (Diptera, Syrphidae). Црни квадрати показују логаритамске вредности граничне вероватноће података $\text{Pr}(X/K)$ за различит број кластера (средња вредност за пет поновљених циклуса). Сиви квадрати означавају ΔK . Извор: модификовано из Francuski и сар. (2020).

Просторна Бајесова кластер анализа (алгоритам имплементиран у програме BAPS, Geneland и Tess) је модел који експлицитно укључује и информације о локацији јединки у простору. За разлику од програма Structure, програми BAPS и Geneland започињу анализу одређивањем оптималног броја кластера, а потом јединке додељују у ове кластере. Програм BAPS обухвата пет варијанти генетичког модела са мешањем (*admixture*) које су засноване на различитим сценаријима биолошког узорковања: 1) јединке узорковане из популације без релевантних географских информација; 2) јединке узорковане у низу одабраних географски ограничених подручја; 3) јединке континуирано узорковане из популација за које су познате географске координате; 4) узорковане групе јединки за које се зна да припадају истој деми са познатим географским координатама за групу; 5) узорковање два узорка, једног са познатим генетичким пореклом (основни/полазни узорак) и другог са непознатим (тренутни узорак). Из наведених модела јасно је да програм нуди могућност и непросторне и просторне кластер анализе како на нивоу јединки (модел 1 и 3), тако и група јединки (модел 2 и 4). Одабир одговарајуће варијанте може значајно да утиче на закључивање о популационој структури. Тако, на пример, у случају употребе мање информативних генетичких маркера, географске информације су веома корисне; наиме, применом просторне анализе у генетичком моделу са мешањем, јединке се могу груписати заједно и када маркери нису довољни да то ураде на одговарајући начин. У случају кластеровања група јединки, модел са мешањем анализира добијене фреквенције алела претходно груписаних података и настоји да идентификује оне групе за које постоји довољно статистичких доказа да се учесталости алела између њих разликују (Corander

и сар., 2008а, б). За утврђивање претпостављене везе између јединки или места узорковања на основу њихових локација, алгоритам за просторну анализу у програму BAPS користи процес Воронојеве теселације (*Voronoi tessellation*)³. У циљу графичке визуелизације резултирајућа Воронојева теселација може се повезати са генетичком структуром популације бојењем сваке ћелије у складу са припадношћу јединке одређеном кластеру. Такође, графички приказ може бити и тродимензионалан, где висина површине сваке теселационе ћелије представља непоузданост класификације (висина стубића одговара вредности 1 – постериорна вероватноћа; тако да, равне површине ћелија значе да је вероватноћа додељивања јединке одређеном кластеру близу јединице, а што је стубић виши да је већа и непоузданост груписања дате јединке) (Corander и сар., 2008б).

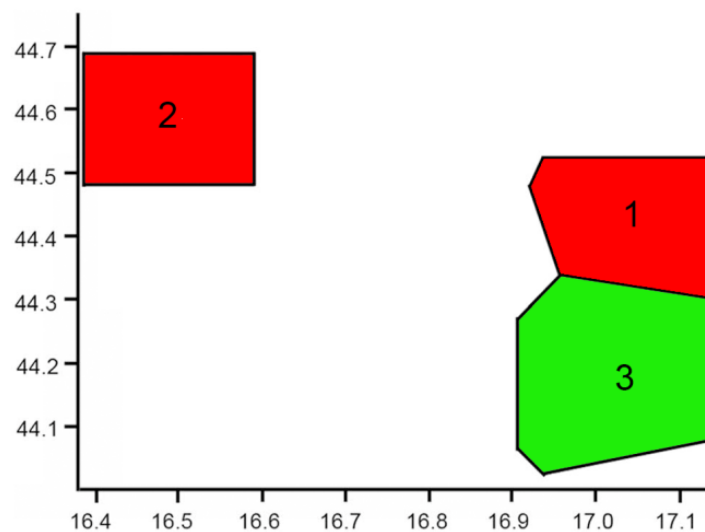
Попут програма BAPS, и програм Geneland имплементира неколико модела који користећи генетичке и географске информације процењују број популација (K) у датом скупу података и утврђују њихову просторну организацију. Процена броја популација се заснива на постериорној вероватноћи која се израчунава помоћу MCMC алгоритма. Програм Geneland такође користи процес Воронојеве теселације и полази од претпоставке да се проучавана географска област сваке популације може представити као скуп региона, полигоналних области, који одговарају просторним обрасцима насталим услед ограниченог протока гена због присуства физичких баријера (на пример, пут, реке, планински ланци и слично) (Guillot и сар., 2005). Просторна дистрибуција претпостављених популација (кластера) може се графички приказати мапом подручја узорковања на којој су бојом означене релативне постериорне вероватноће припадања јединки сваком од кластера.

Пример 4.

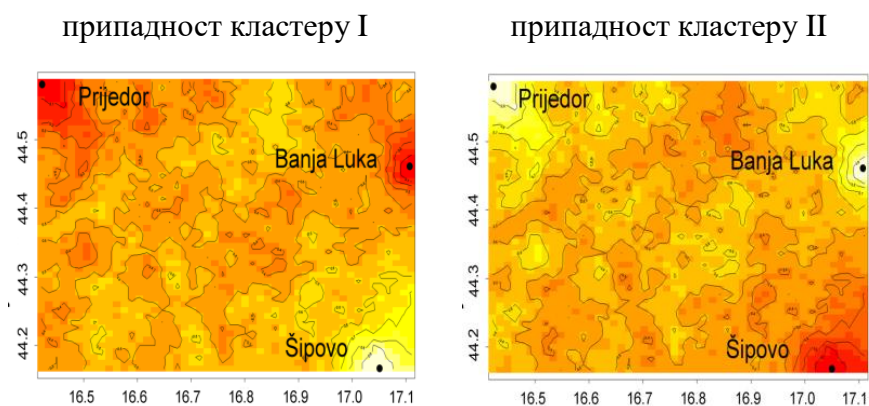
Просторна структурираност генетичког диверзитета врсте *Eristalis arbustorum* (Diptera, Syrphidae) је анализирана на основу алелних фреквенција алозимских локуса (Francuski и сар., 2020). На основу индивидуалних генетичких разлика, а узимајући у обзир и координате локалитета сакупљених јединки, у програму BAPS је утврђена вероватноћа од 98,1% да постоје два генетичка кластера у области истраживања. Један кластер чине јединке узорковане на локалитетима Бања Лука и Приједор, док узорак из Пљеве представља засебну генетичку јединицу (Слика 39). Такође, и просторно експлицитна Бајесова кластер анализа изведена у програму Geneland подржава постојање два јединствена генетичка кластера; кластер I чине Бања Лука и Приједор, а кластер II Пљевља (= Шипово) (Слика 40) (Лукач, 2014; Francuski и сар., 2020). Иако је Бајесова кластер анализа у програму Structure указала на висок степен мешања популација (јединке узорковане на различитим локалитетима чине јединствени генетички кластер; видети раније), резултати анализе геореференцираних генетичких

³ За више детаља о Воронојевој теселацији видети у делу 10.4. Детектовање потенцијалних баријера за проток гена

података (BAPS и Geneland) су у сагласности са параметрима генетичке диференцијације (F_{ST} параметар и AMOVA) који су такође указали на структурираност популације. Наиме, статистички значајна F_{ST} вредност утврђена је за парове узорака Пљева/Бања Лука ($F_{ST} = 0,109$; $P = 0,011$) и Пљева/Приједор ($F_{ST} = 0,157$; $P = 0,008$). AMOVA је показала да је већи део укупне варијабилности дистрибуиран унутар популација (86,75%), али да варијабилност између популација, иако мања (13,25%), потврђује генетичку диференцијацију између популација ($F_{ST} = 0,132$; $P < 0,001$) (Francuski и сар., 2020).



Слика 39. Генетичко груписање јединки *Eristalis arbustorum* (Diptera, Syrphidae) добијено методом просторне Бајесове кластер анализе у програму BAPS. Различите боје означавају различите кластере. Идентична боја указује на популације са хомогеним генетичким саставом. 1 - Бања Лука, 2 - Приједор, 3 - Пљева. Извор: слику уступила Милица Лукач.



Слика 40. Мапа подручја узорковања са релативним постериорним вероватноћама припадности јединки *Eristalis arbustorum* (Diptera, Syrphidae) генетичким кластерима (I и II) добијена програмом Geneland. Црне тачке представљају географски положај узоркованих јединки (Шипово = Пљева). Бела боја означава високу (100%), а црвена ниску (0%) вероватноћу. Извор: слику уступила Милица Лукач.

Бајесов приступ, на моделу заснован алгоритам за процену генетичке структуре популације на основу ДНК секвенци или других података молекуларних маркера, имплементиран је у многе бесплатно доступне софтвере који се међусобно разликују у експлицитним претпоставкама модела, усвојеним стратегијама и детаљима израчунавања на којима се базира статистичко закључивање. Међу њима се као најчешће коришћени издвајају:

- Програм Structure (Pritchard и сар., 2000; Falush и сар., 2003, 2007) је софтверски пакет намењен за проучавање популационе структуре, а на основу мултилокусних генотипских података. Може се применити на већину најчешће коришћених генетичких маркера, укључујући микросателите, SNP, RFLP и AFLP. На основу разлика у фреквенцији алела Structure може да идентификује присуство различитих популација (генетичких кластера), додели јединке тим кластерима на основу анализе вероватноће, идентификује мигранте или јединке чији су преци били имигранти, утврди удео у геному јединке који потиче из различитих претпостављених популација (*admixture analysis*), процењује генетичку дистанцу између претпостављених популација и предачких популација.
- BAPS (*Bayesian analysis of population structure*; Corander и сар., 2003, 2008a) је програм за Бајесову анализу генетичке структуре популације. Може се применити на фреквенције алела молекуларних маркера (или нуклеотидне фреквенције за податке о ДНК секвенци), али и број генетички дивергентних

група у популацији као случајне варијабле. Анализе и поређења модела се такође могу извршити коришћењем фиксног броја генетички дивергентних група или унапред дефинисане структуре популација. Применом програма може се извршити груписање молекуларних података и анализа генетичког мешања (*admixture analysis*) на индивидуалном и групном нивоу применом непросторног и просторног приступа. Експлицитно просторни приступ захтева да анализирани јединке буду геореференциране (тј. да буду познате географске координате места узорковања сваке јединке).

- Програм Geneland (Guillot и сар., 2005) је доступан као R пакет. На основу геореференцираних индивидуалних мутилокусних генотипова омогућава закључивање о броју популација и просторној локацији генетичких дисконтинуитета између тих популација. Такође, израчунава и параметре F статистике између претпостављених популација.
- Програм Tess (Chen и сар., 2007; Durand и сар., 2009) намењен је за процену и визуелизацију генетичке структуре популације на основу мутилокусних генотипова јединки узоркованих на различитим географским локацијама, а без претпоставке о унапред дефинисаним популацијама. Програм имплементира алгоритме за просторну анализу за моделе са и без генетичког мешања (*admixture* и *without admixture*) и нарочито је погодан за проналажење генетичких дисконтинуитета у континуираним популацијама и процену просторног варирања пропорција генетичког мешања (*admixture*). Може се користити и за претраживање генома и раздвајање адаптивне од неадаптивне генетичке варијабилности на основу диференцијације у фреквенцији предачких алела (Caue и сар., 2016).

10.2.3. Методе просторне аутокорељације

Геофизички процеси и коришћење земљишта од стране човека у различите сврхе (пољопривредне, стамбене, индустријске, комерцијалне и друге) условљавају просторну структурираност срединских фактора на које организми реагују, док биотички процеси попут просторно ограниченог парења и дисперзије јединки додатно утичу на просторну структуру популација (Wagner и Fortin, 2016). Зато је проучавање односа просторне структуре генетичких и предеоних података од посебног значаја у предеоној генетици. За процену повезаности генетичке сродности и географске удаљености парова јединки користе се методе просторне аутокорељације (*spatial autocorrelation*). Уопштено говорећи, аутокорељација је статистичка процедура којом се мери просечна сличност било која два посматрања у зависности од њихове просторне, временске или филогенетске удаљености. Просторна аутокорељација је метода просторне статистике којом се у предеоној генетици утврђује степен до ког су јединке које су ближе једна другој у простору генетички сличније од оних које су удаљеније (Balkenhol и сар., 2016). Тестира се да ли утврђени генотип јединке на једном

локалитету зависи од генотипа јединке на суседном локалитету (Manel и сар., 2003). Просторна аутокорејација се обично квантификује Морановим параметром I чија вредност може варирати од -1 до 1. Позитивна вредност указује на позитивну просторну аутокорејацију, што значи да јединке које су просторно ближе су сличније од оних удаљених. Вредност параметра $I = 0$ указује на независност, односно одсуство значајне аутокорејације, док негативна вредност значи да су међусобно ближи узорци различитији од оних удаљених (негативна аутокорејација) (Wagner и Fortin, 2016). Графичким приказивањем вредности коефицијента аутокорејације (Моранов параметар I) у односу на генетичку удаљеност формира се корелограм (тј. график вредности корелејације) који омогућава идентификовање просторних генетичких образаца, те закључивање о потенцијалним процесима (на пример, изолација дистанцом, изолација баријером, клина и слично) који су утицали на утврђену генетичку структуру. Ипак, иако се просторном аутокорејацијом може одредити размера просторног обрасца (глобална или локална структура), она не може да идентификује специфичну локацију генетичког дисконтинуитета у пределу (тј. баријере за проток гена попут река, планина и других) (Manel и сар., 2003).

10.3. Визуелизација просторних образаца дистрибуције генетичког диверзитета: метода интерполације

Процена генетичке структуре традиционално заснована на статистичким параметрима за парове популација (попут F_{ST} параметра или мера генетичке дистанце) не омогућава сагледавање утицаја свих популација истовремено. Боље разумевање просторних образаца генетичког диверзитета се може постићи визуелизацијом података добијених аналитичким методама у виду графичког приказа. Наиме, образац генетичких дистанци између анализираних узорака у пределу се графички може приказати као тродимензионална површина која се назива генетички облик предела (*genetic landscape shape*; Miller, 2005). На овај начин, уместо да се само идентификују суседне области узорковања између којих су вредности генетичке дистанце велике, омогућава се квалитативна карактеризација свих области анализираних географског предела (Miller, 2005). Дакле, формира се приказ површине предела на коме x и y осе одговарају географским локацијама узорковања, а z оса (висина површине) представља генетичку дистанцу. Добијена слика је претпостављени приказ образаца диверзитета у датом пределу који садржи врхове у оним подручјима где су генетичке дистанце велике. Другим речима, конструисе се генетички предео индивидуалних генотипова чији врхови и долине осликавају присуство или одсуство генетичке диференцијације између популација. Конструсање генетичког облика предела се заснова на процесу интерполације. Овим процесом се врши процена података на неузоркованим локацијама у пределу (интерполиране вредности) помоћу математичког модела просторног обрасца узоркованих вредности. Развијено је више типова модела

интерполације, укључујући моделе инверзне пондерисане дистанце, кригинга (*kriging*), теселације и друге (Storfer и сар., 2007).

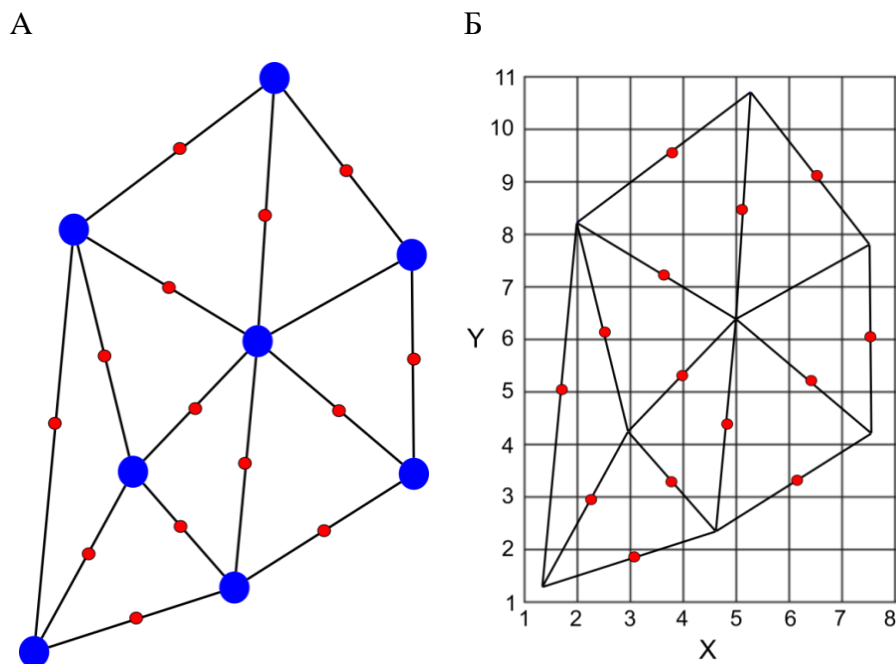
Модели интерполације:

- **Модел инверзне пондерисане дистанце** (*inverse distance weighting*) се заснива на претпоставци да су тачке које су ближе једна другој сличније од оних које су удаљеније. За предвиђање вредности на неузоркованим локацијама, користе се измерене вредности које окружују дато место за које се интерполацијом предвиђају вредности. Измереним вредностима које су најближе локацији за коју се интерполира даје се већа тежина (утицај, значај; *weight*) у односу на оне даље лоциране (Watson и Philips 1985; Watson, 1992).
- **Кригинг** (*kriging*), познат и као Гаусов регресиони процес, користи познате вредности на узоркованим локацијама и параметре из семивариограма за процену вредности на неузоркованим локацијама. Семивариограм графички приказује просторну аутокорељацију (зависност) између измерених тачака узорка. Када је сваки пар локација уцртан, модел се уклапа кроз њих. Вредности семивариограма постају веће како запажања постају мање слична, дајући на тај начин процену удаљености између тачака пре него што оне постану некорелисане (Storfer и сар., 2007).

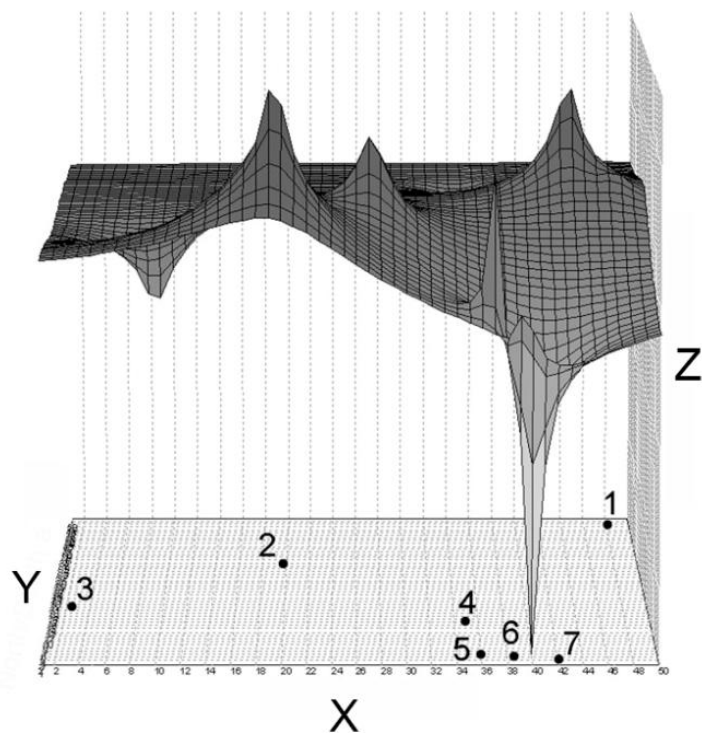
Генерисање приказа површине предела започиње формирањем мреже локалитета на којима су јединке узорковане, тако да међусобно повезани локалитети (плави кружући) формирају троуглове чија темена представљају координате локалитета (Слика 41А). Два су могућа начина за формирање мреже: први, да се сви локалитети међусобно повежу (тј. сваки локалитет са свим осталим), и други, да се повезивање локалитета у мрежи заснива на примени Делонијеве триангулације (Слика 41А; *Delaunay triangulation*; Watson, 1992, Brouns и сар., 2003). Када се анализа заснива на просторним подацима за релативно мали број тачака узорковања први начин може бити примењив. Међутим, то у многим истраживањима није случај, па се зато знатно чешће користи други начин (Miller, 2005).

Након формирања мреже, за све парове популација повезаних локалитета се израчунавају генетичке дистанце и постављају на средину сваке везе (црвени кружићи; Слика 41А). Ове вредности генетичке дистанце ће одредити висину површине (вредности на z оси). У сврху интерполације, за израчунавање висине површине могу се користити сирове генетичке дистанце, резидуали генетичких дистанци добијени линеарном регресијом генетичких дистанци на географске удаљености (користе се када у скупу података постоји корелација између генетичких и географских дистанци; Manni и сар., 2004), или “псеудонагиби” (*pseudoslopes*; кориговане генетичке дистанце добијене за сваки пар популација као количник њихове генетичке и географске дистанце) (Miller, 2005).

Преко формиране мреже повезаних локалитета поставља се правоугаона мрежа одређених димензија (мрежа има одређени број редова и колона, димензије се могу мењати; Слика 41Б) и процедуром интерполације утврђују висина површине за све чворове (место пресека хоризонталних и вертикалних линија) дате мреже. Дакле, на основу вредности генетичких дистанци између стварних тачака узорковања (плави кружићи; Слика 41А) и претпостављених (интерполираних) вредности израчунатих за све координате на правоугаоној мрежи формира се тродимензионална површина предела (Слика 42). На Слици 42 приказан је генетички предео конструисан методом интерполације на основу фреквенције алела алозимских локуса јединки седам европских популација врсте *Eristalis tenax* (јединке су сакупљене у Финској, Немачкој, Швајцарској и Србији) (Francuski и сар., 2013). На приказаном генетичком пределу уочавају се географска подручја повећане и смањене међуиндивидуалне генетичке удаљености (врхови и удолине указују на високе и ниске генетичке удаљености између јединки). Високи пикови указују на генетичку диференцијацију између различитих популација, чак и оних географски блиско лоцираних (Francuski и сар., 2013). Интерпретацијом изгледа генетичког предела у контексту географских одлика истраживаног подручја могу се детектовати карактеристике предела које ограничавају или олакшавају проток гена.



Слика 41. Мрежа узоркованих локалитета (плави кружићи) формирана применом Делонијеве триангулације (А) и правоугаона мрежа димензија 7 x 10 поља (Б). Црвени кружићи означавају вредности генетичке дистанце израчунате за јединке парова повезаних локалитета.



Слика 42. Приказ површине генетичког предела врсте *Eristalis tenax* (Diptera, Syrphidae) добијен методом интерполације у програму Alleles in space. X и Y осе одговарају географским локацијама узорковања јединки анализираних популација, а Z оса (висина површине) представља генетичку дистанцу. Црним кружићима је означен географски положај места узорковања јединки: 1 - Финска; 2 - Немачка; 3 - Швајцарска; 4, 5, 6, 7 - Србија. Извор: модификовано из Francuski и сар. (2013).

Програм Alleles in space (AIS; Miller, 2005) обухвата низ статистичких процедура које су засноване искључиво на анализи међуиндивидуалних образаца генетичке и географске варијабилности намењених за откривање или карактеризацију образаца просторне генетичке структуре. AIS је погодан за анализу генетичких података када је узоркована једна или неколико јединки на великом броју локација у проучаваном пределу. Програм је дизајниран да анализира различите типове генетичких података, укључујући кодоминантне генотипове (алозими, микросателити), доминантне маркера (AFLP, RFLP) и ДНК секвенце. Анализе укључене у AIS: анализа просторне аутокореелационе за откривање генетичке структуре и закључивање о просторним скалама на којима се генетичка структура јавља, анализа индекса агрегације алела за карактеризацију просторне дистрибуције јединки у пределу, Мантелов тест за утврђивање корелације између генетичких и географских дистанци јединки,

детектовање локације претпостављених баријера за проток гена применом Монмонијеровог алгоритма, генерисање генетичког облика предела методом интерполације за визуелизовање образаца генетичких удаљености између узорака у пределу (Miller, 2005).

10.4. Детектовање потенцијалних баријера за проток гена

Све карактеристике географских предела (геоморфолошке и еколошке одлике, елементи људске активности) утичу на проток гена, при чему неке структуре (на пример, путеви, водени токови или планински гробен) могу бити потенцијално непремостиве баријере за одређене популације (Storfer и сар., 2007). Генетичким подацима се могу идентификовати нагли прекиди у протоку гена, као и постепени прелази. За идентификовање генетичких граница (у просторним анализама, граница је област на мапи генетичке варијабилности где долази до наглих промена) користе се методе сврстане у две главне групе: 1) методе Бајесове кластер анализе (*Bayesian clustering methods*); и 2) методе детекције ивица (*edge detection methods*) (Safner и сар., 2011).

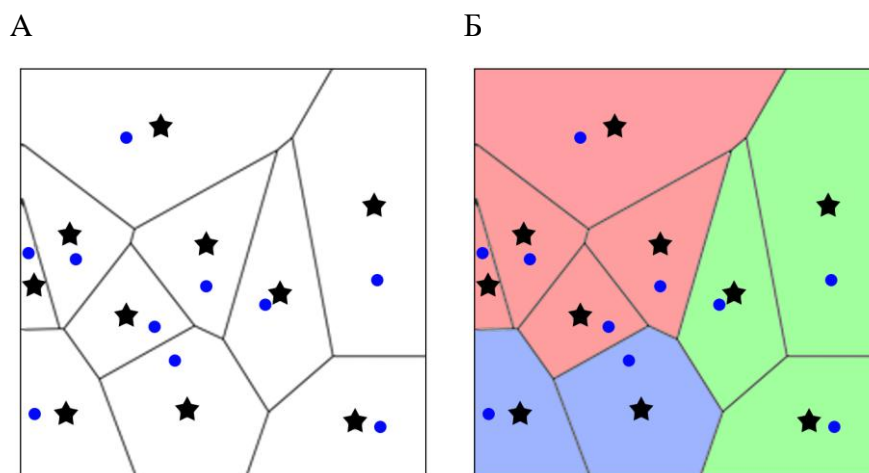
10.4.1. Детектовање баријера у просторној Бајесовој кластер анализи

Као што је раније већ речено⁴, методе Бајесове кластер анализе идентификују највероватније дискретне скупове генетички сродних јединки (кластере) на основу њихових мултилокусних генотипова. Јединка се разврстава (додељује) у кластер за који је њена постериорна вероватноћа највећа. Припадност јединки одређеним кластерима може се користити за препознавање потенцијалних генетичких дисконтинуитета у пределу. Међутим, да би потенцијална баријера (једна или више њих) за проток гена између препознатих кластера јединки била детектована, неопходно је да анализирани генотипови буду геореференцирани, односно да географске координате узоркованих локалитета буду укључене у кластер анализу. Наиме, у просторним моделима, географска удаљеност између две јединке утиче на вероватноћу да оне припадају истом кластеру.

Дакле, просторни модели Бајесове кластер анализе се ослањају на географски експлицитне претходне дистрибуције (*prior distribution*). У моделима инкорпорираним у најзначајније програмске пакете попут Geneland, BAPS и Tess претходна дистрибуција се заснива на процесу Воронојеве теселације (*Voronoi tessellation*), мада неки програми користе и скривени Марковљев модел (*hidden Markov random fields*).

⁴ Видети део 10.2.2. Методе Бајесове кластер анализе

Воронојева теселација представља статистички модел који дефинише суседство сваке јединке теселацијом, односно поделом равни проучаваног простора на непреклапајуће области назване Воронојеве ћелије или полигони (Слика 43А) (François и сар., 2006; Corander и сар., 2008б). Наиме, за проучавани географски регион, сваку узорковану јединку карактеришу податак о њеној просторној локацији (дводимензионалне просторне координате) и податак о генетичким маркерима (генотип). Сматра се да је у датом географском простору присутно K различитих популација и да свака популација заузима одређени регион, део тог простора. Делећи географски простор на ове регионе, алгоритам процењује вредност K (број популација, генетичких кластера) и фреквенције алела. Претпоставља се да сваки регион приближно одговара скупу непреклапајућих конвексних полигона (= Воронојевих ћелија). Пошто не знамо где ови полигони треба да буду постављени, модел користи случајан број тачака које су расуте по пределу са униформном дистрибуцијом (униформна дистрибуција подразумева да свака вредност у скупу могућих вредности има исту вероватноћу да се деси). Свака од ових тачака представља језгро (*nucleus*) конвексног полигона (Слика 43А). Тачке узорковања јединки које се налазе унутар одређеног полигона припадају том језгру. Два места узорковања су суседи ако њихове ћелије деле заједничку ивицу. Процес формирања полигона путем случајног броја и постављања тачака (језгара) назива се и Поасон-Воронојева теселација, јер се случајни број језгара генерише Поасоновим процесом (стохастички процес у којем се догађаји дешавају непрекидно и независно један од другог). Под претпоставком да све популације имају једнаку претходну вероватноћу, онда сваки полигон у географском простору има вероватноћу $1/K$ да припада одређеној популацији (Warwick и Beerli, 2015). Полигони се додељују различитим просторним кластерима, а процена вредности K (оптималан број просторних кластера/популација) и појединачне вероватноће додељивања добијају се коришћењем МСМС алгоритма (Марковљев ланац Монте Карло) (Safner и сар., 2011). Вероватноћа да било која два полигона припадају истом просторном кластеру опада са географским растојањем између њих (Guillot и сар., 2005). Потенцијалне баријере за проток гена се затим повлаче дуж ивица полигона, тако да генотипови на једној страни полигона припадају једном кластеру, а генотипови на другој страни припадају другом (Guillot и сар., 2005). Приликом графичког приказивања полигони који припадају истом кластеру биће обојени истом бојом (Слика 43Б).



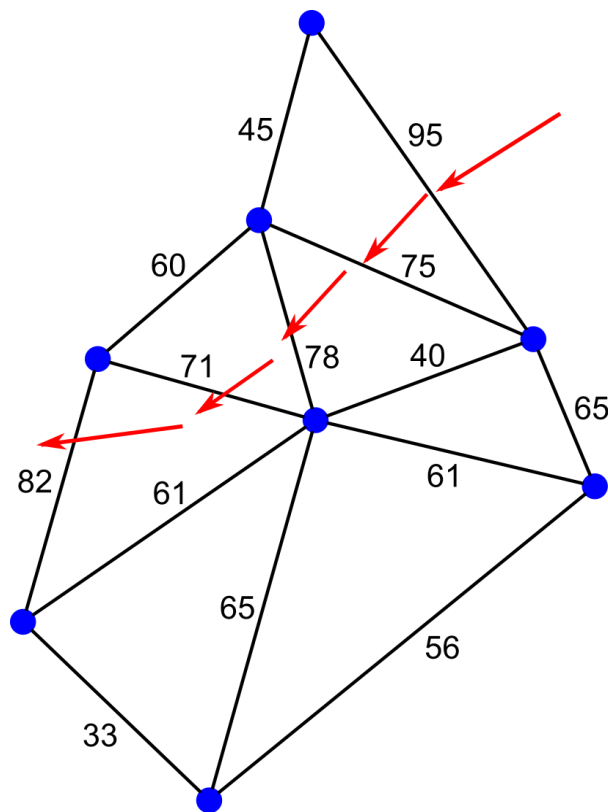
Слика 43. Пример Поасон-Воронојеве теселација. А) Воронојеве ћелије су раздвојене црним линијама. Положај језгра (*nucleus*) сваког полигона (ћелије) приказан је звездicom, а место узорковања плавим кружићем. Б) Припадност географских локација различитим кластерима приказано је различитим бојама. У датом примеру $K = 3$.

10.4.2. Детектовање баријера методама детекције ивица

Монмонијеров алгоритам (*Monmonier's maximum difference algorithm*; Monmonier, 1973) и вомблинг (*wombling*; Womble, 1951) су методе које на основу података о просторној дистрибуцији фреквенције алела идентификују генетичке дисконтинуитете (генетичке границе), односно позицију претпостављене баријере за проток гена (Safner и сар., 2011). Ове методе се могу означити и као директне методе детекције баријера/граница, јер не подразумевају претходно препознавање генетичких кластера јединки, већ се утврђују границе између различитих узорака.

Монмонијеров алгоритам се заснива на мрежи локалитета узоркованих популација или јединки која је конструисана процесом триангулације (повезивање позиција суседних узорака) и израчунавању генетичке дистанце за сваку ивицу (линија која спаја суседне узорке) у мрежи (Слика 44). Положај узорака у мрежи одговара њиховој релативној географској позицији. На основу матрице генетичких дистанци, алгоритам препознаје постојање баријере (једну или више њих) које издвајају популације (или групе популација) између којих је проток гена смањен, или је у потпуности одсутан (Слика 44). Наиме, алгоритам прво проналази највећу вредност генетичке дистанце између било која два међусобно повезана локалитета на мрежи и то представља почетни сегмент баријере. Потом, продужава баријеру на ивицу непосредних суседа са највећом генетичком дистанцом и наставља даље све док се не

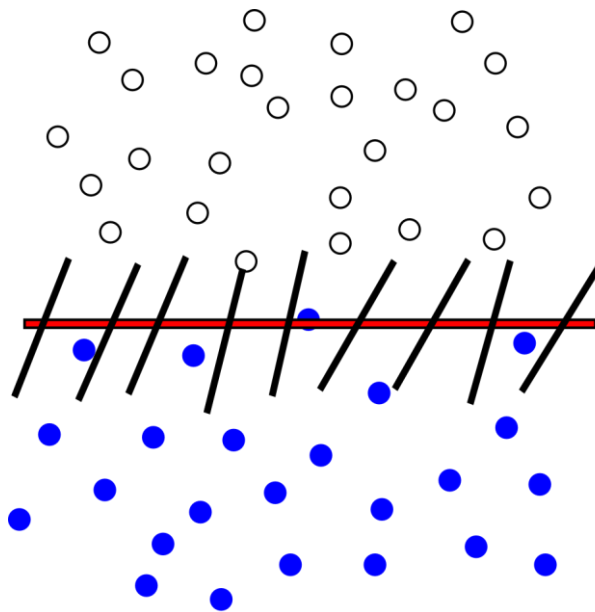
постигне гранична вредност (за дефинисање граничне вредности може се употребити графички приказ вредности генетичких дистанци поређаних по опадајућој вредности, гранична вредност ће вероватно довести до наглог смањења ових вредности), или док растућа баријера не стигне до ивице мреже или унутрашњег сегмента који је претходно дефинисан као сегмент баријере. Дакле, на овај начин се проналази путања која показује највеће генетичке дистанце између повезаних локалитета. Продужавање баријере се врши у правцу веће генетичке удаљености између локалитета, а уколико се утврде идентичне вредности генетичке дистанце прате се оба правца (Слика 44) (Manel и сар., 2003; Miller, 2005).



Слика 44. Пример хипотетичке мреже локалитета (плави кружићи) узоркованих популација/јединки конструисане Делонијевом триангулацијом. Вредности генетичке дистанце између два узорка дате су на свакој ивици (ивица је приказана црном линијом која спаја два географски суседна локалитета) формираних троуглова. Положај и правац пружања баријере означен је црвеним стрелицама.

Други приступ за откривање региона на мапи генетичке варијабилност у којима долази до наглих промена, назван вомблинг, користи површину добијену интерполацијом фреквенција алела и на њој открива положај баријере тако што проналази регионе где је апсолутна вредност нагиба површине висока. Дакле, прво се вредности континуиране варијабле, односно фреквенције алела трансформишу у

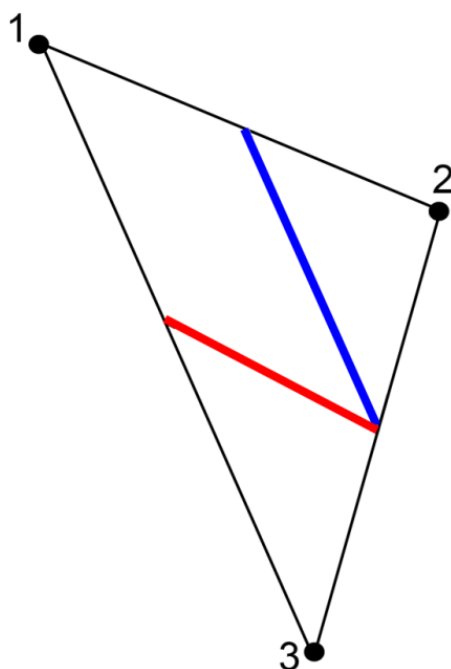
континуирану дистрибуцију процесом интерполације и/или конструкцијом правилне решетке у чијим чворовима се налазе узорци (површина/решетка се формира за сваки алел). Потом се за фреквенције алела израчунава парцијални извод у свакој тачки решетке (или површини), бележећи степен и правац максималног нагиба. За све површине фреквенција алела израчунава се просечна вредност величине и угла нагиба у свакој тачки решетке и добија системска функција у којој је свака тачка решетке представљена вектором који има своју величину и правац. Региони површине који имају високе вредности функције су управо оне области где је стопа промене фреквенција алела највећа и стога се могу идентификовати границе или баријере (Слика 45) (Vocquet-Appel и Vasro, 1994; Manel и сар., 2003; Safner и сар., 2011).



Слика 45. Пример идентификовања баријере применом вомблинг методе на основу фреквенција алела једног локуса јединки узоркованих на различитим локалитетима (бели и плави кружићи). Црне линије означавају степен и правац нагиба површине (дужина линије је пропорционална величини нагиба), односно правац и величину промене фреквенције алела. Црвена линија означава баријеру. Извор: модификовано из Manel и сар. (2003).

Пример 5.

Применом Монмониеровог алгоритма идентификоване су потенцијалне препреке за дисперзију јединки врсте *Eristalis arbustorum* (Diptera, Syrphidae) унутар подручја истраживања (Босна и Херцеговина; места узорковања јединки су Бања Лука, Приједор и Пљева). Визуелизација просторног распореда хаплотипова *COI* мтДНК открила је географску баријеру за проток гена која раздваја популацију из Бања Луке у односу на популације узорковане у Приједору и Пљеви (Слика 46). С друге стране, дистрибуција генотипова алозимских локуса сугерише да баријера одваја јединке из Пљеве у односу на преостала два узорка (Слика 46). Иако употребљени генетички маркери нису сагласни у положају потенцијалне баријере за проток гена, анализе просторне дистрибуције генетичког диверзитета потврђују присуство генетичког дисконтинуитета између узорака (Francuski и сар., 2020).



Слика 46. Мрежа географских локација узоркованих јединки врсте *Eristalis arbustorum* (Diptera, Syrphidae) и положај генетичке баријере одређен применом Монмонијеровог алгоритма у програму Alleles in space на основу алозимских података (црвена линија) и на основу *COI* мтДНК података (плава линија). 1 - Бања Лука, 2 - Приједор, 3 - Пљева. Извор: модификовано из Francuski и сар. (2020).

11. Корелисаност генетичког, географског и фенотипског дисконтинуитета

Након идентификовања просторно-генетичких образаца у популацији може се тестирати њихова повезаност (корелација) са срединским варијаблама, одликама предела, или морфолошким карактеристикама. Наиме, квантификовање генетичке диференцијације између популација помоћу генетичких мера (попут F_{ST} параметра, генетичке дистанце и других) и статистичких метода (на пример, применом анализе главних компоненти или Бајесовог приступа) не пружа увид у ефекат географског простора на утврђене обрасце дистрибуције генетичке варијабилности унутар и између популација. Међутим, познато је да, на пример, географска удаљеност између популација може условити пропорционално мању стопу миграције, па да се и F_{ST} вредности између парова популација разликују. Тако да, уколико се утврди да је генетичка удаљеност између популација у позитивној корелацији са географском дистанцом између њих, може се претпоставити да је просторна дистрибуција генетичке варијабилности резултат процеса изолације дистанцом (*isolation-by-distance*; Wright, 1969). И не само удаљеност између популација (обично мерена праволинијски ваздушном линијом између две локације), већ и различите компоненте предела (планине, реке и слично) могу утицати на кретање јединки, те стога и креирати различите обрасце дистрибуције генетичке варијабилности код различитих врста (Manel и сар., 2003; Storfer и сар., 2007).

Мантелов тест је најчешће коришћена метода за процену односа између географске дистанце и генетичке дивергенције (Mantel, 1967; Diniz-Filho и сар., 2013). Тест има широку примену у биологији, укључујући предеону екологију и генетику, за утврђивање просторне структурираности у подацима или детектовање просторне корелације између два скупа података (матрице). Данас је Мантелов тест у својим различитим облицима (Мантелов тест, парцијални Мантелов тест, корелограм) имплементиран у више компјутерских програма и R пакета намењених за популационе генетичке анализе (на пример, Alleles in space, Arlequin, пакет *vegan* и други) (Diniz-Filho и сар., 2013). Осим повезаности генетичке дистанце са географском дистанцом, Мантеловим тестом се може тестирати и однос генетичке дистанце са различитим срединским варијаблама, као и морфолошком варијабилности.

Мантелова статистика првобитно је заснована на једноставном векторском производу:

$$Z = \sum_{i=1}^n \sum_{j=1}^n x_{ij} \times y_{ij}$$

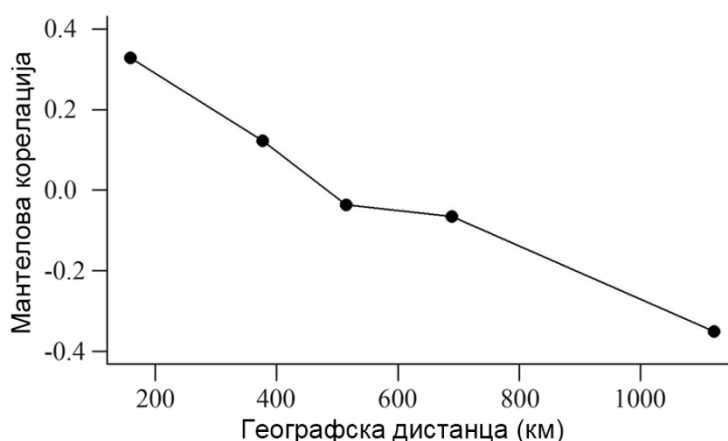
где x и y представљају дистанце (генетичка, географска, или нека друга) између i -те и j -те популације из скупа од n анализираних популација (Mantel, 1967). Међутим, пошто је овако израчуната Z вредност зависна од броја популација и величине њихових дистанци, изведена је и стандардизована верзија теста која користи средње вредности и

варијансе X и Y матрица дистанце. Овај облик Мантеловог теста заправо представља Пирсонову корелацију (r ; *Pearson's correlation*) између стандардизованих елемената две матрице, па тако Z вредност близу 0 указује на одсуство корелације између матрица и, обрнуто, вредности близу -1 и 1 да повезаност постоји и ког је смера. Стандардизовани Мантелов тест се назива и Мантелова корелација (Diniz-Filho и сар., 2013). Нулта хипотеза Мантеловог теста је одсуство корелације између вредности две матрице дистанце, а значајност коефицијента корелације (r) две матрице се утврђује пермутационим приступом.

Парцијални Мантелов тест је облик Мантеловог теста којим се, такође, утврђује корелација између две матрице генетичких или фенотипских дистанци, при чему се у обзир узима и ефекат треће матрице (обично је то матрица географске удаљености). Наиме, парцијалним Мантеловим тестом се израчунава корелација између резидуала две главне матрице дистанце након линеарне регресије сваке матрице на трећу матрицу дистанце (Diniz-Filho и сар., 2013; Guillot и Rousset, 2013)⁵.

Мантелов тест (Мантелова корелација) нам показује какав је однос између две матрице. Међутим, бољи увид у однос између генетичких и географских дистанци у простору, нарочито ако овај однос није линеаран може се добити помоћу Мантеловог корелограма (*Mantel correlogram*; Слика 47). Корелограм представља приказ вредности корелације генетичке удаљености у одређеним интервалима географске удаљености. За конструисање Мантеловог корелограма матрица географских дистанци се подели на више подматрица, од којих свака описује парове популација унутар граница интервала географске удаљености. Дефинисање броја и опсега интервала географске дистанце је произвољно и зависи од просторног распореда анализираних популација. График добијен приказивањем вредности корелације за различите интервале географске дистанце омогућава сагледавање варијабилности коефицијента корелације у простору. Тако, на пример, континуирано и линеарно смањење вредности корелације (а тиме и смањење генетичке сличности) са повећањем географске дистанце указује на клинални образац дистрибуције генетичке варијабилности у анализираном простору. Са друге стране, експоненцијално смањење вредности корелације (високе вредности Мантелове корелације у првим интервалима географске дистанце, које потом опадају и стабилизују се након одређене дистанце) указује да у датом простору постоје острва (*patches*) генетичке варијабилности или сличности узрокованих изолацијом дистанцом, присуством баријере или неким другим фактором (Слика 47; Diniz-Filho и сар., 2013; Cushman и сар., 2016).

⁵ О регресионој анализи видети у делу 5.3. Мултиваријантне статистичке методе.

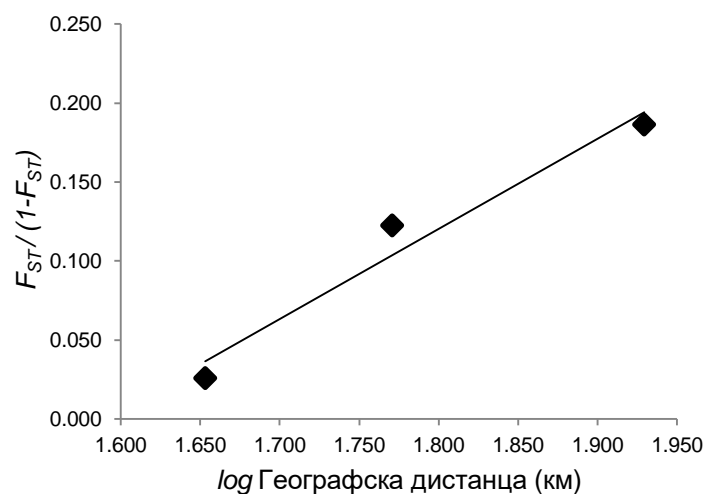


Слика 47. Мантелов корелограма. Извор: модификовано из Diniz-Filho и сар. (2013).

Мантеловим тестом и парцијалним Мантеловим тестом се не може квантификовати количина варијабилности објашњена срединским факторима. Зато се као алтернативна метода може користити каноничка коресподентна анализа (*Canonical correspondence analysis*; ССА) која омогућава сагледавање везе генетичког диверзитета и фактора животне средине, те идентификовање оних срединских варијабли које значајно доприносе разумевању варијабилности генетичког диверзитета (Manel и сар., 2003; François и Waits, 2016). ССА подразумева унимодални однос приликом описивања градијената у првом сету варијабли (зависна варијабла; на пример, фреквенције алела популација) у погледу другог сета варијабли (независна варијабла). Осим ССА, такође се могу применити и редундантна анализа (*Redundancy analysis*) или просторна анализа еигенфункције (*Spatial eigenfunction analysis*). Ове алтернативне методе заснивају се на трансформацији географског простора у форму сирових података (на пример, координате или индекси повезаности) уместо у виду матрице дистанце (Diniz-Filho и сар., 2013; François и Waits, 2016).

Пример 6.

У раније поменутом истраживању генетичке структуре врсте *Eristalis arbustorum* (Francuski и сар., 2020) применом Мантеловог теста између матрица генетичке дистанце (за парове популација израчунате су вредности $F_{ST}/(1 - F_{ST})$; Rousset 1997) и географске дистанце (логаритамски трансформисане вредности удаљености између локалитета узорковања популација) није утврђена статистички значајна корелација ($r = 0,9795$, $P = 0,17$), што указује да изолација дистанцом није условила генетичку диференцијацију популација (Слика 48).



Слика 48. Мантелов тест између матрица генетичке и географске дистанце популација врсте *Eristalis arbustorum* (Diptera, Syrphidae). Вредности географске удаљености између места узорковања популација су логаритамски трансформисане.

ПРИЛОГ

ПРИЛОГ 1.

Корак по корак упутство за програме *tps* серије: *tpsUtil*

Програм *tpsUtil* (Rohlf, 2015) садржи бројне функције за манипулисање датотекама, укључујући креирање, промену формата, поделу или комбиновање датотека, брисање или промену редоследа тачака и јединки и друго.

1. Креирање *tps* датотеке (*.tps)

1.1. Све слике (фотографије, снимци или било који други вид дигиталног записа објеката истраживања) сачувати у једном фолдеру;

1.2. Отворити *tpsUtil* и из падајуће листе операција одабрати операцију *Build tps file from images*;

1.3. Кликом на команду *Input* отвара се прозор за одабир фолдера у коме се налазе слике, довољно је селектовати једну од слика и кликнути *Open*;

1.4. Кликом на команду *Output* отвара се прозор за одабир фолдера у коме се налазе слике за које се креира *.tps датотека, потребно је дати име за *tps* датотеку и сачувати кликом на *Save*; (слике и креирана *tps* датотека морају бити у истом фолдеру);

1.5. Кликом на команду *Setup* отвара се прозор са листом слика у одабраном фолдеру у коме се може извршити селекција слика које желимо да буду обухваћене *tps* датотеком;

1.6. Са кликом на команду *Create* потврдити креирање датотеке. Креирана датотека ће садржати поље „LM = 0” и поље „IMAGE = ” за сваку слику одабрану да буде укључена у *tps* датотеку:

LM=0

IMAGE=slika01.jpg

LM=0

IMAGE=slika02.jpg

LM=0

IMAGE=slika03.jpg

LM=0

IMAGE=slika04.jpg

LM=0

IMAGE=slika05.jpg

...

2. Трансформисање контуре/криве у координате специфичних тачака

2.1. Покренути програм tpsUtil и из падајуће листе операција одабрати операцију *Append tps Curve to landmarks*;

2.2. Кликом на команду *Input* учитати сачувану *.tps датотеку са координатама тачака дуж криве/контуре;

2.3. Кликом на команду *Output* доделити ново име за датотеку; име треба да буде различито од имена учитане датотеке како не би прекопирали оригиналну датотеку. Извршити операцију кликом на команду *Create*.

3. Креирање датотеке за дефинисање специфичних тачака и полутачака (*sliders file*) и повезаних специфичних тачака (*link file*)

3.1. Покренути програм tpsUtil и из падајуће листе операција одабрати операцију *Make sliders file*;

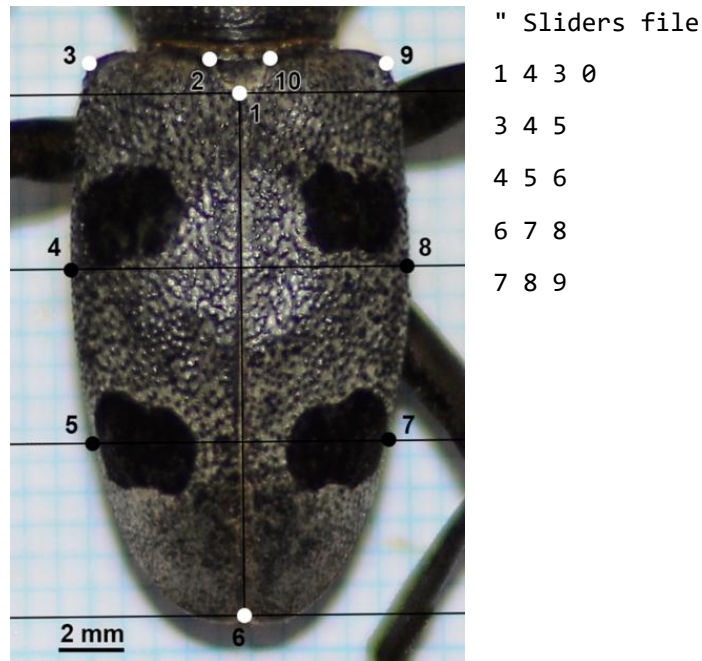
3.2. Кликом на команду *Input* учитати сачувану *.tps датотеку са координатама специфичних тачака. Уколико датотека садржи координате тачака криве/контуре неопходно је прво извршити трансформисање криве/контуре у координате специфичних тачака помоћу операције *Append tps Curve to landmarks*;

3.3. Кликом на команду *Output* доделити име за датотеку која ће бити креирана;

3.4. Кликом на команду *Setup* отвара се прозор са приказаним позицијама специфичних тачака. Помоћу миша повезати по три тачке тако што се кликне на прву тачку и држећи леви тастер миш помера до друге, а потом и треће тачке, те пусти тастер. Тачка у средини ће променити боју (црвени круг постаје бели/празан круг) и тако бити обележена да је полутачка (*semilandmark*). Приликом повезивања тачака, појавиће се линија између њих, а звучни сигнал ће се чути када курсор буде довољно близу друге, односно треће тачке. Повезивање по три тачке наставити док све жељене тачке не буду обележити као полутачке. Десним кликом миша на исцртану везу између тачака отвара се прозор са опцијом за брисање те везе.

3.5. Кликом на команду *Create* датотека са дефинисаним специфичним тачкама и полутачкама (*sliders file*) ће бити сачуван са именом које је претходно додељено у *nts* формату (*.nts). Изглед датотеке у случају када од 10 постављених тачака, тачке 1, 2, 3, 6, 9 и 10 представљају специфичне тачке (бели кружићи), а тачке 4, 5, 7 и 8 полутачке (црни кружићи) (Слика П1): у првом реду се налази име датотеке, други ред је линија са параметрима која садржи четири броја (први и последњи су фиксни за овај програм) од којих други број (у наведеном примеру је број 4) означава да су дефинисане четири полутачке а трећи број (број 3 у примеру) да су подаци форматирани у три колоне. Од трећег до последњег реда датотека у сваком реду садржи низ од три броја где број у

средњој колони означава тачку дефинисану као полутачка, а прва и трећа колона означавају бројеве тачака између којих ће током процеса суперимпозиције бити дозвољено да дата полутачка клизи до достизања оптималног поравнања са консензус конфигурацијом. Редослед по којем су редови наведени није битан и не утиче на процес суперимпозиције.



Слика П1. Положај специфичних тачака (бели кружићи) и полутачака (црни кружићи) и изглед датотеке којом се дефинишу полутачке. Извор: модификовано из Gojković и сар. (2022).

Такође, програм `tpsUtil` даје могућност да се дефинише веза између било којег пара специфичних тачака помоћу операције *Make link file*. Као и у претходно описаном случају креирања датотеке (*sliders file*), након читавања датотеке са координатама специфичних тачака и именовања датотеке која ће бити креирана (видети кораке 3.2. и 3.3.) кликом на команду *Setup* отвара се прозор са приказаним позицијама специфичних тачака. Помоћу миша повезати две тачке тако што се кликне на прву тачку и држећи леви тастер миш помера до друге, те пусти тастер (звучни сигнал ће се чути када курсор буде довољно близу друге тачке). Између две повезане тачке појавиће се линија. Десним кликом миша на исцртану везу отвара се прозор са опцијом за брисање те везе. Кликом на команду *Create* датотека са повезаним специфичним тачкама (*link file*) ће бити сачуван са именом које је претходно додељено у *nts* формату (*.nts). Креирана датотека садржи у првом реду име датотеке, у другом реду су параметри (четири броја од којих су први и последњи фиксни за овај програм) од којих други број (у наведеном примеру је број 3) означава да су дефинисане везе, а трећи број (број 2 у примеру) да су подаци форматирани у две колоне. Од трећег до последњег реда датотека у сваком реду

садржи низ од два броја означавајући бројеве специфичних тачака које су повезане. Редослед по којем су редови наведени није битан и не утиче на процес суперимпозиције. Ако на примеру приказаном на слици П1 направимо везу између специфичних тачака 3, 6 и 9, датотека ће изгледати:

```
" Sliders file
```

```
1 3 2 0
```

```
3 9
```

```
3 6
```

```
6 9
```

Осим на претходно описани начин који пружа могућност да се графички дефинишу везе између одређених тачака, датотеке (*sliders file* и *link file*) се такође могу припремити и у неком од доступних текст процесора (на пример, Notepad) и сачувати у облику *.nts.

Корак по корак упутство за програме *tps* серије: *tpsDig*

Програм *tpsDig* (Rohlf, 2015) је намењен за прикупљање морфометријских података у форми координата специфичних тачака (*landmarks*), полутачака (*semilandmarks*), крива и контура, као и линераних мера, углова, обима и површина. Специфичне тачке и једна или више контура/крива могу се заједно поставити на једној слици, а њихове координате комбиновати за даљу анализу уз претходно превођење координата тачака криве/контуре у координате специфичних тачака. Прикупљање координата захтева поштовање неколико основних правила:

- * На свакој слици мора се поставити исти број специфичних тачака;
- * Редослед специфичних тачака на свакој слици мора бити исти;
- * За места постављања специфичних тачака одабрати кореспондирајуће (хомологне) тачке, односно тачке које могу бити јасно идентификоване на свим сликама на основу заједничке морфологије, функције, или уобичајене геометрије;
- * У случају да се у узорку налазе слике објеката које имају различиту оријентацију (на пример, слике десног и левог крила), слике је потребно окренути тако да све буду окренуте у исту страну.

1. Постављање специфичних тачака (*landmarks*)

1.1. Покренути програм tpsDig;

1.2. Из падајућег менија *File* са опцијом *Input source* ► *File...* отворити креирану *.tps датотеку (Напомена: све слике и *tps* датотека морају се налазити у истом фолдеру);

1.3. Кликом активирати алатку за постављање специфичних тачака (иконица *Digitize landmarks*; изгледа као нишан) и левим кликом миша на одређено место на слици објекта поставити специфичне тачке;

Ако су објекти истраживања фотографисани са мерном скалом, пре постављања специфичних тачака за сваку слику може се подесити размера (тј. одредити фактор скале):

1.3.1. У падајућем менију *Options* отворити *Image tools* (или кликом на иконицу *Image edit tools*) и отварањем прозора овог менија изабрати картицу *Measure*;

1.3.2. У отвореном прозору у пољу *Reference length* дефинисати познату дужину и јединицу мере;

1.3.3. Кликнути на команду *Set scale*, па на слици на мерној скали једном кликнути мишем на почетку мере, померити курсор и потом једном кликнути на крају дужине дефинисане мере; кликом на команду *OK* потврдити и затворити прозор;

1.3.4. Фактор размере израчунат за прву слику биће примењен и на све наредне слике у узорку уколико се не изврши подешавање размере за сваку слику посебно;

1.3.5. Израчунати фактор размере биће сачуван заједно са позицијама координата тачака у *tps* датотеци (поље "SCALE = "). Напомена: координате специфичних тачака се увек чувају у *tps* датотеци у пиксел јединицама које одговарају слици;

1.4. Десним кликом миша алатка за постављање тачака се мења у алатку за едитовање (*Edit landmarks & curves*; бела стрелица ↗), постављањем стрелице на место специфичне тачке бела стрелица се мења у црну стрелицу ↕ помоћу које се тачке могу померати (са кликом и држањем левог тастера, те померањем миша) или брисати (десним кликом миша на тачку се отвара прозор са опцијом за брисање специфичне тачке);

1.5. Иконице ← и → (црвене стрелице у лево и десно) служе за прелазак на следећу слику, или повратак на претходну;

1.6. Иконице *Fit*, + и - служе за увећавање и смањивање приказа на екрану;

1.7. У падајућем менију *Options* чекирањем опције *Label landmarks* поред постављених специфичних тачака ће се приказати и њихови редни бројеви;

1.8. Опција *Image tools* (у падајућем менију *Options* или кликом на иконицу *Image edit tools*) осим за одређивање фактора скале (подешавање размере; видети кораке 1.3.1. - 1.3.5.) служи и за подешавање величине и боје за специфичне тачке и бројеве (картица *Color*), окретање, сечење и конвертовање слика (картица *Tools*), одабир изгледа курсора алатке за постављање специфичних тачака (картица *Cursors*);

1.9. По завршетку постављања специфичних тачака, *tps* датотеку сачувати са именом под којим је датотека првобитно креирана користећи из падајућег менија *File* опцију *Save data ...* и *Overwrite existing file*. Сачувана *tps* датотека ће за сваку слику садржати поља „LM = ” са бројем постављених тачака, *x* и *y* координате свих постављених специфичних тачака, поље „IMAGE = ” са називом слике и поље “ID = “ са идентификационом ознаком (програм ће за сваку слику редом у датотеци аутоматски додељивати нумеричку ознаку почев од нуле, а истраживач може те ознаке преименовати). Уколико је за слике дефинисан и фактор размере, налазиће се и поље “SCALE = “:

LM=14

3321.00000 1446.00000
3223.00000 1467.00000
3157.00000 1466.00000
3141.00000 1404.00000
3015.00000 1432.00000
3044.00000 1462.00000
2823.00000 1390.00000
2652.00000 1381.00000
2364.00000 1412.00000
2360.00000 1447.00000
2371.00000 1479.00000
2690.00000 1514.00000
2800.00000 1547.00000
2698.00000 1459.00000

IMAGE=slika01.jpg

ID=0

SCALE=0.056453

LM=14

3259.00000 1349.00000
3171.00000 1348.00000
3118.00000 1348.00000
3113.00000 1289.00000
2992.00000 1298.00000
3019.00000 1325.00000
2817.00000 1245.00000
2641.00000 1249.00000
2413.00000 1312.00000
2413.00000 1346.00000
2423.00000 1377.00000
2700.00000 1369.00000
2779.00000 1402.00000
2718.00000 1309.00000

IMAGE=slika02.jpg
ID=1
SCALE=0.061699
...

2. Креирање криве/контуре

2.1. Покренути програм tpsDig;

2.2. Из падајућег менија *File* са опцијом *Input source* ► *File...* отворити креирану *.tps датотеку (Напомена: све слике и tps датотека морају се налазити у истом фолдеру);

2.3. Кликом активирати алатку за цртање криве/контуре (иконица *Draw background curves*; жута оловка);

2.4. У падајућем менију *Options* штриклирати (✓) опцију *Close curve* за цртање затворене контуре (*outline*); за цртање отворене криве није потребно укључити ову опцију;

2.5. На слици исцртати криву/контур у на следећи начин: левим кликом миша од јасно дефинисаног места започети цртање, померајући миш у смеру кретања казавки на сату дуж контуре наставити са кликтањем (са сваким кликом линија ће бити све дужа);

2.6. Десним кликом миша алатка за цртање криве/контуре се мења у алатку за едитовање (*Edit landmarks & curves*; бела стрелица ↗), померањем стрелице преко линије криве/контуре на линији ће се појавити тачке. Померањем беле стрелице на тачку, она се мења у црну стрелицу ↖ помоћу које се може променити позиција тачке (са кликом и држањем левог тастера, те померањем миша) а тиме и линија контуре/криве, или обрисати тачка (десним кликом миша на тачку се отвара прозор са опцијом за брисање специфичне тачке). Померањем беле стрелице на линију између две тачке, она се мења у симбол + , те се левим кликом тада може додати нова тачка на линију;

2.7. Десним кликом миша отвара се мени, одабиром опције *Resample curve* отвара се прозор који приказује тренутни број тачака на криви/контур и омогућава одређивање новог броја тачака; може се дефинисати жељени број тачака по критеријуму “*By steps*” (и тада нови број мора бити мањи од постојећег) или “*By length*” (задати број тачака ће бити распоређен на једнакој међусобној удаљености дуж криве/контуре);

2.8. Мени који се у *Edit modu* (тј. када је активна бела стрелица) отвара на десни клик миша садржи и опције за израчунавање дужине криве/контуре (*Curve length*) и површине контуре (*Curve area*), а информација ће бити исписана испод прозора (*Status panel* на дну екрана). За криве, површина се израчунава тако што алгоритам креира затворену површину повлачењем линије између прве и последње тачке криве. Уколико је пре цртања криве/контуре било извршено скалирање, приказане вредности ће бити

изражене у јединици дефинисаној при скалирању, у супротном ће вредности бити изражене у пикселима;

2.9. По завршетку цртања криве/контуре, *tps* датотеку сачувати са именом под којим је датотека првобитно креирана користећи из падајућег менија *File* опцију *Save data ...* и *Overwrite existing file*. Сачувана *tps* датотека ће за сваку слику садржати поља „LM = ” са бројем постављених тачака, поље “CURVES = ” са бројем исцртаних крива/контура, поље “POINTS = “ са бројем тачака у свакој од нацртаних крива/контура, *x* и *y* координате свих постављених тачака, поље „IMAGE = ” са називом слике и поље “ID = “ које садржи идентификациону ознаку. Уколико је за слике дефинисан и фактор размере, налазиће се и поље “SCALE = “. Изглед датотеке када су постављене четири специфичне тачке и једна крива која има 10 међусобно једнако удаљених тачака дуж ње:

```
LM=4
234.00000 835.00000
1670.00000 734.00000
1673.00000 600.00000
185.00000 598.00000
CURVES=1
POINTS=10
233.00000 838.00000
266.00000 831.00000
299.00000 827.00000
332.00000 824.00000
365.00000 821.00000
399.00000 820.00000
432.00000 818.00000
466.00000 816.00000
499.00000 816.00000
532.00000 818.00000
IMAGE=slika01.jpg
ID=0
LM=4
260.00000 835.00000
1543.00000 804.00000
1538.00000 694.00000
246.00000 498.00000
CURVES=1
POINTS=10
256.00000 837.00000
286.00000 832.00000
317.00000 832.00000
347.00000 828.00000
377.00000 828.00000
408.00000 828.00000
438.00000 827.00000
469.00000 827.00000
499.00000 827.00000
```

```
530.00000 828.00000
IMAGE=slika02.jpg
ID=1
...
```

2.10. Отворити програм tpsUtil и помоћу операције *Append tps Curve to landmarks* трансформисати координате тачака дуж криве/контуре у координате специфичних тачака. Нова сачувана *tps* датотека ће за сваку слику садржати поље „LM = ” са бројем постављених тачака (укупан број тачака), *x* и *y* координате свих постављених тачака, поље „IMAGE = ” са називом слике и поље “ID = “ са идентификационом ознаком. Уколико је за слике дефинисан и фактор размере, налазиће се и поље “SCALE = “. Изглед датотеке из корака 2.9. након трансформисања криве/контуре у специфичне тачке:

```
LM=14
234.00000 835.00000
1670.00000 734.00000
1673.00000 600.00000
185.00000 598.00000
233.00000 838.00000
266.00000 831.00000
299.00000 827.00000
332.00000 824.00000
365.00000 821.00000
399.00000 820.00000
432.00000 818.00000
466.00000 816.00000
499.00000 816.00000
532.00000 818.00000
IMAGE=slika01.jpg
ID=0
LM=14
260.00000 835.00000
1543.00000 804.00000
1538.00000 694.00000
246.00000 498.00000
256.00000 837.00000
286.00000 832.00000
317.00000 832.00000
347.00000 828.00000
377.00000 828.00000
408.00000 828.00000
438.00000 827.00000
469.00000 827.00000
499.00000 827.00000
530.00000 828.00000
IMAGE=slika02.jpg
ID=1
...
```

3. Мерење линеарних дистанци и углова

3.1. Покренути програм tpsDig;

3.2. Из падајућег менија *File* са опцијом *Input source* ► *File...* отворити креирану *.tps датотеку (Напомена: све слике и *tps* датотека морају се налазити у истом фолдеру);

3.3. Кликом активирати алатку за мерење дистанци (иконица *Make linear measurements*; изгледа као лењир) или из падајућег менија *Modes* одабрати операцију *Measure mode*;

3.4. Левим кликом миша обележити почетак, затим померити стрелицу курсора на крај линеарне мере и кликнути поново; на крају исцртане линије биће приказана вредност дужине мере у пикселима (ако није извршено скалирање) или јединици дефинисаној приликом скалирања;

3.5. Кликом активирати алатку за мерење углова (иконица *Make angle measurements*; нацртан је угао) или из падајућег менија *Modes* одабрати операцију *Angle mode*;

3.6. Са три клика миша дефинисати угао, а вредност измерена у степенима ће се појавити уз сам угао;

3.7. На једној слици може се измерити више линеарних мера; десним кликом отвара се прозор са опцијама *Add to listing...* и *Add all to listing...* да се поједине или све измерене вредности додају на листу;

3.8. У падајућем менију *File* одабиром опције *View listing...* отвара се прозор у коме се налазе информације о улазној датотеци и измереним вредностима. Све информације из овог прозора могу се сачувати, копирати или штампати.

ПРИЛОГ 2.

Корак по корак упутство за програме *tps* серије: *tpsRelw*

Програм *tpsRelw* (Rohlf, 2015) је намењен за статистичку анализу варијабилности облика у узорку. Изводи суперимпозицију конфигурација специфичних тачака и полутачака и анализу главних компоненти варијабилности облика. Програм даје могућност графичког приказивања усаглашене (просечна, консензус) конфигурације специфичних тачака и положаја тачака појединачних конфигурација у односу на консензус, као и дистрибуције варијабилности облика у морфолошком простору помоћу графика расејања (*scatterplot*), те промене облика дуж оса главних компоненти.

1. Суперимпозиција конфигурација специфичних тачака

1.1. Покренути програм *tpsRelw*;

1.2. У падајућем менију *Options* проверити подешавања (*Alfa = 0*; *Uniform component: complement*; *Scale aligned: 1*; *Projection: orthogonal*; *Alignment: Procrustes*). Напомена: ако је одабрана опција *Alignment: Procrustes* скалиране координате поравнатих јединки (Прокрустове координате) неће садржати центроидну величину, а уколико нам је потребно да сачувамо координате поравнатих јединки и задржимо информацију о центроидној величини потребно је одабрати опцију *Alignment: Boas coordinates* (то су Прокрустове координате за сваку јединку помножене њеном оригиналном центроидном величином);

1.3. У делу прозора *Input* кликом на команду *Data* учитати *.tps датотеку са сачуваним координатама специфичних тачака. Уколико датотека садржи криву/контуру, неопходно је да тачке криве/контуре претходно буду трансформисане у координате специфичних тачака помоћу операције *Append tps Curve to landmarks* у програму *tpsUtil* (видети Прилог 1);

1.4. У делу прозора *Compute* мишем кликнути на команде *Consensus*, *Partial warps* и *Relative warps* (тим редоследом);

1.5. У делу прозора *Display* може се добити графички приказ извршених операција;

1.5.1. Кликком на *Consensus* приказује се усаглашена конфигурација специфичних тачака (*consensus*). У падајућем менију *Options* одабиром опције *Points* и/или *Vectors* биће приказане позиције специфичних тачака и вектори одступања специфичних тачака од усаглашеног облика за све јединке у узорку;

1.5.2. Кликком на *Partial warps* приказује се график матрице *partial warp* скорова (*W* матрица). Позиција сваке јединке је одређена *x* и *y* компонентом прве *partial warp* (могу се одабрати и друге *partial warp* за приказ);

1.5.3. Кликком на *Relative warps* приказује се график матрице *relative warp* скорова; позиција сваке јединке је одређена паром *relative warp* (може се селектовати одређена комбинација *relative warp* за приказ);

1.6. У падајућем менију *File* опција *Save...* отвара понуду низа различитих матрица које се могу сачувати како би се могле користити у другим програмима. На пример, датотека *Weight matrix* представља матрицу *partial warp* скорова која као “матрица података” може бити улазна датотека за мултиваријантну анализу облика у стандардним статистичким пакетима, а датотека *Centroid size* је матрица која садржи информацију о центроидној величини за сваку јединку у узорку.

Датотеке различитих матрица могу се сачувати у неком од понуђених формата (*tps*, *nts*, *csv*). У наставку је дат пример изгледа датотеке центроидне величине у **.nts* формату за узорак од 15 јединки, а на основу положаја десет специфичних тачака (вредности центроидне величине су исказане у пикселима, јер није извршено скалирање). У првом реду је назив датотеке, у другом реду су параметри који показују колико јединки је обухваћено узорком (15L у наведеном примеру) и да су подаци форматирани у једну колону (трећи број у низу: 1), у трећем реду је наведен низ идентификационих ознака јединки обухваћених узорком, од четвртог реда надаље налазе се вредности центроидне величине за јединке према редоследу наведеном у низу идентификационих ознака.

```
" Centroid size
1 15L 1 0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
2.38966517298349E+002
1.85838919947033E+002
2.30268459831604E+002
2.62239520538955E+002
2.48087431316275E+002
2.41415086128672E+002
1.91645351190734E+002
1.55809196476353E+002
2.13338967228255E+002
1.72946830883374E+002
1.82204158436037E+002
2.02996964730042E+002
2.04233736985431E+002
2.74853483333831E+002
2.24390098115820E+002
```

За наведени пример у наставку је дат изглед датотеке *W* матрице (матрица *partial warp* скорова). У првом реду је назив датотеке, у другом реду су параметри који показују колико јединки је обухваћено узорком (15L у наведеном примеру) и да су подаци форматирани у 16 колона (трећи број у низу: 16L), у трећем реду је наведен низ идентификационих ознака јединки обухваћених узорком, од четвртог реда надаље налазе се вредности скорова 16 варијабли облика за сваку од јединки. Редослед јединки

одговара редоследу наведеном у низу идентификационих ознака. Напомена: током процеса суперимпозиције конфигурација специфичних тачака долази до смањења димензионалности простора облика (по једна димензија се губи приликом рескалирања конфигурација на јединицу центроидне величине и ротације, а две приликом translације конфигурација), тако да ће дводимензионална конфигурација са k специфичних тачака након суперимпозиције бити описана са $2k-4$ варијабли (у наведеном примеру $k = 10$, а број варијабли 16).

```
" Weight matrix (rows are specimens, cols are x,y pairs)
1 15L 16L 0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
1X 1Y 2X 2Y 3X 3Y 4X 4Y 5X 5Y 6X 6Y 7X 7Y UniX UniY
-1.78019288125116E-003 -9.55551091899571E-004 -5.86955990447569E-003
-2.54680495437740E-003 3.74454632696588E-003 -1.28076492858173E-003
-2.61364934104940E-003 -7.50271112699719E-003 -4.82569615005914E-005
-2.86279085653280E-003 -1.39778153472883E-002 1.92378127606988E-004
3.19897849159863E-003 -4.56786589039783E-003 -1.33251257290908E-002
2.17443165106491E-003 7.68757694615025E-004 2.88846312936823E-003
-2.02787261041616E-004 -1.18720875118602E-003 -1.40201470318088E-003
7.63859636192741E-004 1.37746630024386E-003 -7.30329585540875E-003
-2.95833855859592E-003 4.38904128894923E-003 -4.38175180788849E-003
-4.91696073775117E-004 6.35997242149185E-003 -4.16564712391011E-003
-1.14081204433472E-002 -1.68772812605152E-003 2.59248778479868E-003
7.40824106807361E-004 -2.16553542433940E-003 2.42342471116719E-003
-2.29241562441974E-004 -4.51509252777680E-003 -7.59612772142190E-004
2.58856805590527E-003 8.57735646185383E-003 4.67679671820505E-003
8.81632414923508E-003 1.15138895333948E-002 4.65476175250418E-003
4.16547905514201E-003 1.59375475240700E-004 6.71088757515833E-003
...
```

2. Суперимпозиција конфигурација полутачака (*semilandmarks*)

1. Креирати *sliders* фајл у програму tpsUtil (видети Прилог 1); *sliders* фајл дефинише које тачке треба да буду третиране као полутачке;

2.1. Покренути програм tpsRelw;

2.2. У падајућем менију *Options* проверити подешавања (*Alfa = 0*; *Uniform component: complement*; *Scale aligned: 1*; *Projection: orthogonal*; *Alignment: Procrustes*);

2.3. У делу прозора *Input* кликом на команду *Data* учитати *.tps датотеку са сачуваним координатама тачака. Уколико датотека садржи криву/контуру, неопходно је да тачке криве/контуре претходно буду трансформисане у координате специфичних тачака помоћу операције *Append tps Curve to landmarks* у програму tpsUtil (видети Прилог 1);

2.4. У падајућем менију *File* одабрати опцију *Open sliders file...* и учитати претходно креирану датотеку (*sliders file*) у којој је дефинисано које тачке треба третирати као полутачке;

2.5. У падајућем менију *Options* у оквиру опције *Slide method* одабрати жељени критеријум за суперимпозицију полутачака (критеријум *Chord - d^2* дозвољава клизање полутачака између две тачке док се не пронађе најмања могућа Прокрустова дистанца између прилагођеног положаја и одговарајуће тачке у консензус конфигурацији, а критеријум *Chord - BE* док се не пронађе положај у које је потребна најмања могућа енергија савијања за деформацију консензус конфигурације). Више о наведеним критеријумима видети у Perez и сар. (2016);

2.6. У делу прозора *Compute* мишем кликнути на команде *Consensus*, *Partial warps* и *Relative warps* (тим редоследом);

2.7. У делу прозора *Display* може се добити графички приказ извршених операција;

2.8. У падајућем менију *File* опција *Save...* отвара понуду низа различитих матрица које се могу сачувати како би се могле користити у другим програмима.

ПРИЛОГ 3.

Корак по корак упутство за програме *tps* серије: *tpsRegr*

Програм *tpsRegr* (Rohlf, 2015) се користи за истраживање односа између облика и једне или више независних променљивих (на пример, величине); обухвата мултиваријантни тест који тестира колико се добро може предвидети варијабилност облика користећи једну или више независних варијабли. У овом програму се могу урадити различите врсте мултиваријантних анализа варијансе (MANOVA) и мултиваријантних анализа коваријансе (MANCOVA), вишеструке регресије за предвиђање независне променљиве у функцији облика, израчунавање Прокрустове дистанце сваког узорка од референтног облика и од сваког узорка до конфигурација предвиђених линеарном регресијом.

1. Мултиваријантна регресија за утврђивање алометријског ефекта

1.1. Покренути програм *tpsRegr*;

1.2. У падајућем менију *Options* проверити почетна подешавања (✓ *PCA align reference*; *Uniform component: Linerized Procrustes*; *Regr. method: multivariate regr.*; ✓ *Do multivariate tests*; *Scale method: 1*; *Projection method: orthogonal*; *Alignment: Procrustes*). Напомена: ако је одабрана опција *Alignment method: Procrustes* скалиране координате поравнатих јединки (Прокрустове координате) неће садржати центроидну величину, а уколико нам је потребно да сачувамо координате поравнатих јединки и задржимо информацију о центроидној величини потребно је одабрати опцију *Alignment method: Boas coordinates* (то су Прокрустове координате за сваку јединку помножене њеном оригиналном центроидном величином);

1.3. У делу прозора *Data* кликом на команду *Data* учитати *.tps датотеку са сачуваним координатама специфичних тачака. Уколико датотека садржи криву/контуре, неопходно је да тачке криве/контуре претходно буду трансформисане у координате специфичних тачака помоћу операције *Append tps Curve to landmarks* у програму *tpsUtil* (видети Прилог 1). Потом кликом на команду *Indep. var.* учитати датотеку са сачуваним центроидним вредностима за јединке анализираног узорка. Напомена: датотека са вредностима центроидне величине за сваку јединку у узорку се мора сачувати у формату *.nts у програму *tpsRelw* (видети Прилог 2);

1.4. У делу прозора *Computation* мишем кликнути на команде *Consensus*, *Partial warps*, *Regression* и *Perm. test* (тим редоследом). Кликом на *Perm. test...* отвара се прозор у коме се може одредити број пермутација, величина блока и метода пермутација. Параметар величине блока и алтернативне методе пермутације (*Within blocks*, *Among blocks*) користе се када су посматрања груписана, као у случају поновљеног тестирања

исте јединке. Тестови пермутације су базирани на параметру Wilks' Λ (ламбда) и F статистици (F -тест).

1.5. У делу прозора *Display* може се добити графички приказ извршених операција;

1.5.1. Кликом на *Consensus* приказује се усаглашена (консензус) конфигурација специфичних тачака. У падајућем менију *Options* одабиром опције *Points* и/или *Vectors* биће приказане позиције специфичних тачака и вектори одступања специфичних тачака од усаглашеног облика за све јединке у узорку;

1.5.2. Кликом на *Shape vs X* приказује се график односа *partial warps* и независне варијабле (центроидна величина). Свака тачка представља јединку анализираног узорка;

1.5.3. Кликом на *Visualization* приказује се деформација (промене) облика повезана са променом независне варијабле. Померањем клизача добијају се предвиђања промене облика за мање или веће вредности независне променљиве;

1.6. У падајућем менију *File* одабиром опције *View report...* отвара се прозор са приказом нумеричких резултата. Биће приказано следеће: информације о улазној датотеци, координате специфичних тачака консензус конфигурације која ће се користити као референтна или тангентна конфигурација, матрица коефицијената регресије (први ред одговара пресеку Y осе и увек ће бити нула, следи табела са квадратом вишеструке корелације за предвиђање сваке компоненте облика у функцији независних променљивих, F вредности, степени слободе и вероватноће), резултати неколико мултиваријантних тестова, Прокрустове дистанце за сваку јединку од консензус конфигурације и проценат варијабилности (*Percent explained*) објашњен независном варијаблом тј. центроидном величином;

1.7. У падајућем менију *File* опција *Save...* отвара понуду низа различитих матрица које се могу сачувати како би се могле користити у другим програмима.

Мултиваријантна регресиона анализа се може употребити и за утврђивање зависности облика и других независних варијабли попут неке еколошке варијабле, географске варијабле, узраста и слично.

2. Мултиваријантна анализа варијансе

С обзиром да се мултиваријантна анализа варијансе (MANOVA) користи за утврђивање статистичке значајности разлика у облику између група, предуслов за њено извођење је да су на самом почетку дефинисане групе јединки. У програму *tpsRegr* се MANOVA изводи мултиваријантном регресијом варијабли облика на независну варијаблу која садржати информацију о груписању јединки. Дакле, пре самог извођења анализе неопходно је припремити датотеку за независну варијаблу у којој ће бити

кодирана припадност јединки различитим групама. Овај тип варијабли се називају “лажне” (*dummy*) варијабле и креира се у виду матрице кодирањем чланова групе помоћу бројева 1, 0 и -1. Ако, на пример, анализирамо три групе јединки матрица ће имати две колоне (број *dummy* варијабли је једнак броју група минус 1). У првој колони сви чланови прве групе обележавају се са 1, сви чланови последње групе са -1, а сви преостали чланови између са 0. У другој колони, са 1 се обележавају сви чланови друге групе, са -1 чланови последње групе, а са 0 сви преостале. Креирана матрица са кодовима се уз заглавље стандардно за датотеке креиране програмима *tps* серије мора сачувати у *.nts формату. У наставку је дат изглед сачуване датотеке за пример са три групе (свака група садржи по пет јединки). У првом реду се налази назив датотеке, у другом реду су параметри матрице од који број 15 означава да матрица има 15 редова (тј. јединки), а 2L да је форматирана у две колоне, у трећем реду се налазе називи колоне (тј. *dummy* варијабли), а у наставку следе кодови:

```
"indep vars
1 15 2L 0
grup1 grupa2
1      0
1      0
1      0
1      0
1      0
0      1
0      1
0      1
0      1
0      1
-1     -1
-1     -1
-1     -1
-1     -1
-1     -1
```

Након што је припремљена датотека, може се урадити MANOVA.

2.1. Покренути програм *tpsRegr*;

2.2. У падајућем менију *Options* проверити почетна подешавања (✓ *PCA align reference*; *Uniform component: Linerized Procrustes*; *Regr. method: multivariate regr.*; ✓ *Do multivariate tests*; *Scale method: 1*; *Projection method: orthogonal*; *Alignment: Procrustes*). Напомена: ако је одабрана опција *Alignment method: Procrustes*, скалиране координате поравнатих јединки (Прокрустове координате) неће садржати центроидну величину, а уколико нам је потребно да сачувамо координате поравнатих јединки и задржимо информацију о центроидној величини потребно је одабрати опцију *Alignment method:*

Boas coordinates (то су Прокрустове координате за сваку јединку помножене њеном оригиналном центроидном величином);

2.3. У делу прозора *Data* кликом на команду *Data* учитати *.tps датотеку са сачуваним координатама специфичних тачака. Уколико датотека садржи криву/контуру, неопходно је да тачке криве/контуре претходно буду трансформисане у координате специфичних тачака помоћу операције *Append tps Curve to landmarks* у програму tpsUtil (видети Прилог 1). Потом кликом на команду *Indep. var.* учитати датотеку са кодовима за груписање јединки анализираних узорка. У падајућем менију *Options* отворити опцију *Select indep. variables...* и у прозору који се отвори са листом варијабли обухваћених датотеком треба селектовати све варијабле.

2.4. У делу прозора *Computation* мишем кликнути на команде *Consensus*, *Partial warps*, *Regression* и *Perm. test* (тим редоследом).

2.5. У падајућем менију *File* одабиром опције *View report...* отвара се прозор са приказом нумеричких резултата неколико мултиваријантних тестова.

ПРИЛОГ 4.

Корак по корак упутство за програм MorphoJ

Интегрисани програмски пакет MorphoJ (Klingenberg, 2011) намењен је за геометријско морфометријску анализу. MorphoJ се може преузети бесплатно на https://morphometrics.uk/MorphoJ_page.html, а упутство за коришћење се налази на https://morphometrics.uk/MorphoJ_guide/frameset.htm?index.htm. Обухвата низ најважнијих мултиваријантних статистичких метода за анализу дводимензионалних и тродимензионалних података облика, укључујући анализу главних компоненти, каноничку и дискриминантну анализу, анализу парцијалних најмањих квадрата, регресију, поређење матрица, процену хипотезе модуларности и процену филогенетског сигнала морфометријских података.

Изглед корисничког интерфејса

Покретањем програма отвара се графички кориснички интерфејс на којем се у врху прозора налазе падајући менији (*Files, Preliminaries...*), а испод њих четири картице (*Project Tree, Results, Graphics* и *Reports*). У прозору картице *Project Tree* се у виду хијерархијски структурираног стабла приказује садржај пројекта и ток анализа, тј. учитани подаци и креиране матрице, те извршене анализе у оквиру датог пројекта. Десним кликом на неку од ставки пројекта отвара се мени у коме су понуђене опције за преименовање и брисање ставке, као и за приказ (текстуални и графички) резултата. У прозору картице *Results* приказују се резултати извршених анализа. Десним кликом отвара се мени у коме су понуђене опције за брисање садржаја у овом прозору (Напомена: брише се само приказ резултата и резултати се могу поново позвати из картице *Project Tree*) и извоз и чување резултата у текстуалном формату (*.txt). У прозору картице *Graphics* налазе се графички прикази резултата распоређених на једној или више картица у оквиру овог прозора. У зависности од извршене анализе графички резултати се приказују у виду промене облика, хистограма, графика расејања (*scatter plot*). Десним кликом отвара се мени са специфичним низом опција (понуђене опције зависе од типа графичког приказа) за додатно подешавање графичког приказа (на пример, одабир једног од четири начина приказивања промене облика, промена оса за приказ или промена боје тачака у графику расејања према одабраној класификацији и слично), као и опције за штампање, те извоз и чување графика у неком од неколико понуђених формата (svg, png, bmp, jpg, gif). И на крају, у прозору картице *Reports* приказују се извештај о извршеним операцијама (статус програма, учитане датотеке, извршене одређене анализе) и грешкама које су се јавиле приликом читавања података или извршавања анализа.

1. Учитавање података и чување датотеке

За креирање новог пројекта из падајућег менија *File* одабиром команде *Create New Project* отвара се прозор у коме је потребно доделити име пројекта и означити да ли су подаци које желимо да учитамо дводимензионални или тродимензионални, да ли је конфигурација специфичних тачака која се анализира билатерално симетрична (тј. има *object* симетрију; оса или раван симетрије пролази кроз конфигурацију специфичних тачака), формат у коме се читавају подаци (одабрати *tps*), а потом означити једну или више *.tps датотека специфичних тачака претходно креираних помоћу *tpsDig* програма. Кликком на команду *Create Dataset* одабрана (или више њих) датотека ће се учитати у програм *MorphoJ* и прозор затворити, а на картици *Project Tree* ће се појавити додељено име пројекта и учитаног сета података. Такође, у већ креирани пројекат може се учитати нови сет података помоћу команде *Create New Dataset* из падајућег менија *File*.

Након што су подаци облика учитани, потребно је додати и категоричке варијабле које ће садржати информације о припадности јединки одређеним групама (на пример, врста, локалитет, пол, узраст, понављање и слично). На картици *Project Tree* селекувати креирани сет података и у падајућем менију *Preliminaries* одабрати команду *Edit Classifiers*. У прозору који се отворио кликом на команду *Add classifier* потребно је доделити име за нову варијаблу и потврдом на команду *OK* појављује се колона у коју потом можемо појединачно за сваку јединку уписати или за све јединке одједном копирати вредности категоричке варијабле из раније припремљене датотеке (најбоље је у *excel* формату припремити све категоричке варијабле за узорак који анализирамо, па појединачне варијабле копирати у *MorphoJ* након додавања нове колоне). Можемо додати једну или више категоричких варијабли и кликом на команду *Accept* потврдити унос. Важно је обратити пажњу да називи група буду конзистентно написани јер је *MorphoJ* осетљив на мала и велика слова, тако да, на пример, ако две јединке припадају истом полу и за једну напишемо „женка“ а за другу „Женка“, програм ће их посматрати као да припадају различитим групама.

Други начин додавања категоричких варијабли је да се из падајућег менија *File* помоћу команде *Import Classifier Variables* прочита раније припремљена текстуална датотека са свим варијаблама које желимо да додамо. Овде је важно означити начин како да се вредности категоричке варијабле придруже одговарајућој јединци у скупу података. Ако се, на пример, одабере да то буде према идентификационој ознаци (*Match by Identifier*), онда се у датотеци са категоричким варијаблама у првој колони морају навести идентификационе ознаке јединки које су идентичне онима из *.tps датотеке конфигурација специфичних тачака.

Креирани пројекат и уређени сетови учитаних података се морају сачувати у формату *.morphoj помоћу команде *Save Project* у падајућем менију *File*. Програм *MorphoJ* ће сачувати пројекат са свим подацима и урађеним анализама, па се командом *Open Project* у падајућем менију *File* може поново отворити и рад наставити у истом пројекту.

2. Суперимпозиција конфигурација специфичних тачака

На картици *Project Tree* селектовати креирани сет података, па из падајућег менија *Preliminaries* одабрати команду *New Procrustes Fit*. У прозору који се отвори одабрати опцију *Align by principal axis* и потврдити кликом на команду *Perform Procrustes fit*. Ако конфигурација специфичних тачака има билатералну симетрију, пре извршавања операције суперимпозиције на картици *Graphics* ће се појавити приказ положаја специфичних тачака са претпоставком које специфичне тачке леве и десне стране структуре чине парове (пар је повезан плавом линијом). Предложено упаривање специфичних тачака се може и кориговати, а потом кликом на *Accept* се прихвата, а суперимпозиција започиње. Након извршене суперимпозиције конфигурација на картици *Graphics* се приказује положај специфичних тачака консензус конфигурације са положајем тачака појединачних конфигурација (када анализирана структура има билатералну симетрију, одвојено се приказују симетрична и асиметрична компонента).

На картици *Project Tree* се селекује сет података за који је урађена суперимпозиција. У падајућем менију *File* се налази команда *Export Dataset* којом се могу у облику *.txt датотеке сачувати оригиналне координате специфичних тачака (*raw coordinates*), центроидна величина (*centroid size*) и Прокрустовете координате (*Procrustes coordinates*) специфичних тачака и потом користи у неком другом програму (за структуре са билатералном симетријом засебно се могу сачувати матрице симетричне и асиметричне компоненте).

3. Анализа главних компоненти (*Principal component analysis*)

На картици *Project Tree* селектовати креирани сет података, па из падајућег менија *Preliminaries* одабрати команду *Generate Covariance Matrices*. У прозору који се отвори *Data type* обележити *Procrustes coordinates* и потврдити на *Execute*. Испод сета података ће се креирана коваријациона матрица појавити као нова ставка (*CovMatrix...Procrustes coordinates*).

На картици *Project Tree* селектовати коваријациону матрицу *CovMatrix...Procrustes coordinates*, потом из падајућег менија *Variation* одабрати команду *Principal Component Analysis*. Испод коваријационе матрице појавиће се две нове ставке (*PCA: CovMatrix...Procrustes coordinates* и *PC scores, CovMatrix...Procrustes coordinates*). Скорови главних компоненти се могу сачувати у виду текстуалне датотеке и користити у неком другом програму тако што се селекује ставка *PC scores* и у падајућем менију *File* одабере опција *Export Dataset*.

На картици *Results* налазе се сви резултати анализе, а десним кликом се отвара мени са опцијом *Export to File* којом се подаци могу сачувати као текстуална датотека (*.txt).

На картици *Graphics* налазе се: 1) графички приказ промена облика за који се десним кликом отвара мени са опцијама за подешавање (на пример, избор осе главних

компоненти за коју се приказују промене, скалирање приказаних промена, тип приказа) и чување графичког приказа (опција *Export Graph to File*); 2) хистограм са приказаним еигенвредностима главних компоненти; 3) график расејања индивидуалних скорова главних компоненти који се може додатно модификовати (на пример, промена оса главних компоненти за приказ, одабир боје тачака) и сачувати.

4. Каноничка анализа (*Canonical variate analysis*)

На картици *Project Tree* селектовати креирани сет података, па из падајућег менија *Comparison* одабрати команду *Canonical Variate Analysis*. У прозору који се отвори одабрати сет података за који желимо да урадимо анализу и тип података (на пример, *Procrustes coordinates*), потом каноничку варијаблу по којој ће подаци бити груписани и штриклирати кућицу за пермутациони тест (подразумевани број пермутација је 10000, али се овај број може променити). Кликом на команду *Execute* биће извршена анализа, а у хијерархијском приказу садржаја пројекта ће се две нове ставке (*CVA...* и *CVA scores...*). Скорови каноничких варијабли се могу сачувати у виду текстуалне датотеке (*.txt) са опцијом *Export Dataset* из падајућег менија *File* уз претходно селектовање ове ставке.

На картици *Results* налазе се сви резултати анализе, а десним кликом се отвара мени са опцијом *Export to File* којом се подаци могу сачувати као текстуална датотека (*.txt).

На картици *Graphics* налазе се: 1) графички приказ промена облика, десним кликом се отвара мени са опцијама за подешавање приказа (на пример, избор каноничке варијабли за коју се приказују промене, скалирање величине приказаних промена, тип приказа) и чување; 2) график расејања индивидуалних скорова каноничких варијабли који се такође може додатно модификовати и сачувати.

5. Дискриминанта анализа (*Discriminant analysis*)

На картици *Project Tree* селектовати креирани сет података, па из падајућег менија *Comparison* одабрати команду *Discriminant Analysis*. У прозору који се отвори одабрати сет података за који желимо да урадимо анализу и тип података (на пример, *Procrustes coordinates*), потом каноничку варијаблу по којој ће подаци бити груписани, само одређене парове или све парове група и штриклирати кућицу за пермутациони тест (подразумевани број пермутација је 10000, али се овај број може променити). Кликом на команду *Execute* биће извршена анализа, а у хијерархијском приказу садржаја пројекта ће се две нове ставке (*Discriminant function...* и *Scores: Discriminant function...*). Дискриминантни скорови се могу сачувати у виду текстуалне датотеке (*.txt) са опцијом *Export Dataset* из падајућег менија *File* уз претходно селектовање ове ставке.

На картици *Results* налазе се сви резултати анализе, а десним кликом се отвара мени са опцијом *Export to File* којом се подаци могу сачувати као текстуална датотека (*.txt).

На картици *Graphics* за сваки анализирани пар група налазе се: 1) графички приказ промена облика који се може додатно модификовати (на пример, скалирање величине приказаних промена, избор типа приказа) и сачувати; 2) хистограм дискриминантних скорова који се десним кликом и одабиром опције *Export Graph to File* може сачувати; 3) хистограм дискриминантних скорова након унакрсне валидације (*cross-validation*) који се такође може сачувати.

6. Мултиваријантна регресија (*Multivariate regression*) варијабли облика на величину - алометрија

На картици *Project Tree* селектовати креирани сет података, па из падајућег менија *Covariation* одабрати команду *Regression*. У прозору који се отвори у левој колони (*Dependent variables*) одабрати сет података, матрицу и варијабле за облик (жељени сет података, Прокрустове координате, Прокрустове координате), а у десној колони (*Independent variables*) за величину (жељени сет података, центроидна величина, центроидна величина или log центроидна величина), потом штриклирати кућицу за пермутациони тест (подразумевани број пермутација је 10000, али се овај број може променити) и кликнути на *Execute*. Резултати анализе ће се као додатна ставка појавити у приказу пројекта, а могу се сачувати у виду текстуалне датотеке (*.txt) са опцијом *Export Dataset* из падајућег менија *File* уз претходно селектовање ове ставке. За сваку јединку у узроку анализа генерише предвиђене вредности и резидуале, као и регресионе скорове што се може користити за даље анализе. У случају статистички значајног ефекта алометрије на облик, анализа облика може се поновити са варијаблима облика коригованим за величину тј. користећи резидуале из регресионе анализе. Тада је на картици *Project Tree* потребно обележити ставку *Regression results*, покренути жељену анализу и од два понуђена типа података (*Predicted values* и *Residuals*) одабрати *Residuals*. (Напомена: за извођење анализе главних компоненти неопходно је прво генерисати коваријациону матрицу одабиром типа података *Residuals*)

На картици *Results* налазе се сви резултати анализе, а десним кликом се отвара мени са опцијом *Export to File* којом се подаци могу сачувати као текстуална датотека (*.txt).

На картици *Graphics* налазе се: 1) графички приказ промена облика, десним кликом се отвара мени са опцијама за подешавање приказа (на пример, скалирање величине приказаних промена, тип приказа) и чување; 2) график расејања индивидуалних регресионих скорова у односу на величину који се, такође, може додатно модификовати и сачувати.

7. Анализа парцијалних најмањих квадрата (*Partial least squares analysis*)

У програму MorphoJ анализа парцијалних најмањих квадрата може се применити за утврђивање обрасца коварирања два одвојена блока (тј. сетова варијабли две анализирани структуре) или између блокова унутар једне конфигурације специфичних тачака.

За анализу коварирања два сета варијабли облика (тј. две структуре) из падајућег менија *Covariation* одабрати команду *Partial Least Squares* и опцију *Two Separate Blocks*, па у прозору који се отвори у левој и десној колони одабрати сетове података, матрице и варијабле облика (жељени сет података, Прокрустове координате, Прокрустове координате), штриклирати кућицу за пермутациони тест (подразумевани број пермутација је 10000, али се овај број може променити) и потом кликом на *Execute* урадити анализу.

За анализу два сета варијабли у оквиру једне анатомске структуре на картици *Project Tree* прво селектовати сет података, а потом из падајућег менија *Covariation* одабрати команду *Partial Least Squares* и опцију *Within a Configuration*. У прозору који се отвори на картици *Select subsets of landmarks* потребно је дефинисати које специфичне тачке из дате конфигурације чине блок 1, а које блок 2. То се може урадити било превлачењем тачака у црвени (сет 1) или плави (сет 2) правоугаоник на дну екрана, или одабиром специфичних тачака из листе са десне стране екрана и додељивањем (*Assign*) одређеном сету (сет 1 или сет 2). Након дефинисања сетова тачака и потврђивања кликом на *Accept* активира се картица *Start analysis* на којој је потребно одабрати тип података (*Procrustes coordinates*) и штриклирати кућицу за пермутациони тест. Кликом на команду *Accept* анализа ће бити урађена, а нова ставка пројекта (*PLS scores*) ће се појавити на картици *Project Tree*.

На картици *Results* налазе се сви резултати анализе, а десним кликом се отвара мени са опцијом *Export to File* којом се подаци могу сачувати као текстуална датотека (*.txt).

На картици *Graphics* налазе се: 1) графички приказ промена облика; 2) хистограм са приказаним уделом сваке од PLS оса у укупној коваријанси; 3) график расејања индивидуалних PLS скорова за два блока, као и парова PLS оса унутар сваког од два блока. За сваки тип графичког приказа резултата се десним кликом отвара мени са опцијама за подешавање приказа и чувања.

8. Друге опције у програму MorphoJ

Јединствени сет података може се поделити на два или више сетова по одабраној категоријској варијабли. На пример, ако смо у јединственом сету анализирали заједно женке и мужјаке, сет можемо поделити на основу припадности одређеном полу и анализе поновити за сваки пол засебно. На картици *Project Tree* обележимо сет података који желимо да поделимо, па из падајућег менија *Preliminaries* одаберемо *Subdivide Dataset By*. Затим одаберемо категоријску варијаблу по којој желимо да буде

извршена подела јединственог сета података (на пример, пол) и да ли да буду креирани нови сетови података за све групе или само за одабране по задатом критеријуму (на пример, само желимо да се креира сет података за женке). Уколико кућица *Separate Procrustes fit* не буде штриклирана, сви нови креирани сетови ће имати заједничку суперимпозицију.

Засебно креирани сетови података у оквиру једног пројекта се могу спојити у нови сет. Из падајућег менија *Preliminaries* одаберемо *Combine Datasets* и одаберемо почетни сет података (*Start dataset*) и сет/сетове које желимо да му придружимо (*Other datasets*).

У падајућем менију *Preliminaries* опција *Include or Exclude Observations* даје могућност да се креира нови сет података који неће садржати одређене јединке/групе, а опција *Selected Landmarks* даје могућност да се креира нови сет података који ће имати другачији редослед специфичних тачака или ће одабране специфичне тачке бити изостављене.

За креирање новог сета података који ће садржати упросечене вредности за јединицу анализе користи се команда *Averages Observations by* из менија *Preliminaries*. На пример, ако је за један узорак три пута извршено постављање специфичних тачака, за сваку јединку из узорка се на основу три забележене конфигурације може израчунати просечна конфигурација специфичних тачака. Неопходан предуслов је да буде додата и категоричка варијабла која ће дефинисати који подаци припадају једној јединици анализе.

Промене облика се графички могу приказати на неколико начина (на пример, мрежом деформације или векторима померања тачака). За визуелизовање промена се може користити и скуп линија које повезују специфичне тачке (*wireframe*). За примену овог графичког приказа неопходно је претходно креирати шеме повезивања тачака. Из падајућег менија *Preliminaries* одабиром команде *Create or Edit Wireframe* којом се може креирати нова или кориговати постојећа шема отвара се приказ конфигурације специфичних тачака. Мишем се повезују по две специфичне тачке док се не повежу све тачке које желимо и потврђује кликом на *Accept*. Креирана шема ће се појавити на картици *Project Tree* као додатна ставка пројекта. На картици *Graphics* се десним кликом на графички приказ промена облика може се променити начин приказивања графика.

ЛИТЕРАТУРА

- Adams, D.C., Collyer, M.L., Kaliontzopoulou, A., Baken, E. (2022) geomorph: Software for geometric morphometric analyses. R package version 4.0.4.
- Adams, D.C., Collyer, M.L. (2016) On the comparison of the strength of morphological integration across morphometric datasets. *Evolution*, 11: 2623-2631.
- Adams, D.C., Otárola-Castillo, E. (2013) geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, 4(4): 393–399.
- Adams, D.C., Rohlf, F.J., Slice, D.E. (2004) Geometric morphometrics: Ten years of progress following the "revolution". *The Italian Journal of Zoology*, 71(9): 5-16.
- Adams, D.C., Rohlf, F.J., Slice, D.E. (2013) A field comes of age: geometric morphometrics in the 21st century. *Hystrix*, 24: 7-14.
- Agarwal, M., Shrivastava, N., Padh, H. (2008) Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports*, 27(4): 617-631.
- Alpert, P., Simms, E.L. (2002) The relative advantages of plasticity and fixity in different environments: when is it good for a plant to adjust? *Evolutionary Ecology*, 16: 285-297.
- Arbour, J.H., Brown, C.M. (2014) Incomplete specimens in geometric morphometric analyses. *Methods in Ecology and Evolution*, 5: 16-26.
- Armbruster, W.S., Pélabon, C., Bolstad, G.H., Hansen, T.F. (2014) Integrated phenotypes: understanding trait covariation in plants and animals. *Philosophical Transactions of the Royal Society B*, 369: 20130245.
- Arnqvist, G., Mårtensson, T. (1998) Measurement error in geometric morphometrics: empirical strategies to assess and reduce its impact on measures of shape. *Acta Zoologica Academiae Scientiarum Hungaricae*, 44: 73-96.
- Balkenhol, N., Cushman, S.A., Storfer, A., Waits, L.P. (2016) Introduction to landscape genetics – concepts, methods, applications. U: Balkenhol, N., Cushman, S.A., Storfer, A.T., Waits, L.P. (Urednici) *Landscape genetics: Concepts, methods, applications*. John Wiley & Sons. pp: 1-7.
- Belkhir, K. (2004) Genetix 4.05.2. University of Montpellier II, Laboratoire Génome et Populations, Montpellier, France.
- Benis, N., Kar, S.K., Martins dos Santos, V.A.P., Smits, M.A., Schokker, D., Suarez-Diez, M. (2017) Multi-level integration of environmentally perturbed internal phenotypes reveals key points of connectivity between. *Frontiers in Physiology*, 8: 388.
- Bocquet-Appel, J.-P., Bacro, J.-N. (1994) Generalized wombling. *Systematic Biology*, 43: 442-448.
- Bogutskaya, N.G., Zupančić, P., Bogut, I., Naseka, A.M. (2012) Two new freshwater fish of the genus *Telestes* (Actinopterygii, Cyprinidae) from karst poljes in Eastern Herzegovina and Dubrovnik littoral (Bosnia and Herzegovina and Croatia). *ZooKeys*, 180: 53-80.
- Bonhomme, V., Picqm, S., Gaucherel, C., Claude, J. (2014) Momocs: Outline Analysis Using R. *Journal of Statistical Software*, 56(13): 1-24.
- Bookstein, F.L. (1991) *Morphometric tools for landmark data: geometry and biology*. Cambridge University Press.

- Bookstein, F.L. (1997) Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape. *Medical Image Analysis*, 1: 225-243.
- Bookstein, F.L., Chernoff, B., Elder, R.L., Humphries, J.M. Jr., Smith, G.R., Strauss, R.E. (1985) *Morphometrics in evolutionary biology*. Academy of Natural Sciences Press, Philadelphia.
- Bookstein, F.L., Gunz, P., Mitteroecker, P., Prossinger, H., Schaefer, K., Seidler, H. (2003) Cranial integration in *Homo*: Singular warps analysis of the mid-sagittal plane in ontogeny and evolution. *Journal of Human Evolution*, 44(2): 167-187.
- Boyer, D.M., Puente, J., Gladman, J.T., Glynn, C., Mukherjee, S., Yapuncich, G.S., Daubechies, I. (2015) A new fully automated approach for aligning and comparing shapes. *Anatomical Records*, 298: 249-276.
- Brouns, G., De Wulf, A., Constales, D. (2003) Delaunay triangulation algorithms useful for multibeam echosounding. *Journal of Surveying Engineering*, 129: 79-84.
- Brun-Usan, M., Zimm, R., Uller, T. (2022) Beyond genotype-phenotype maps: Toward a phenotype-centered perspective on evolution. *BioEssays*, 44: e2100225.
- Byers, D. (2008) Components of phenotypic variance. *Nature Education*, 1(1): 161.
- Campbell, N.A, Atchley, W.R. (1981) The geometry of canonical variate analysis. *Systematic Zoology*, 30: 268-280.
- Cardillo, M. (2010) Some applications of geometric morphometrics to archaeology. U: Elewa, A.M.T. (Urednik) *Morphometrics for non-morphometricians*. Springer. pp: 325-341.
- Caye, K., Deist, T. M., Martins, H., Michel, O., François, O. (2016) TESS3: Fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources*, 16: 540-548.
- CBD (2001) *Global Biodiversity Outlook*. Secretariat of the Convention on Biological Diversity / UNEP: Montreal, Canada
- Chen, C., Durand, E., Forbes, F., François, O. (2007) Bayesian clustering algorithms ascertaining spatial population structure: A new computer program and a comparison study. *Molecular Ecology Notes*, 7: 747-756.
- Chesnokov, Yu.V., Kosolapov, V.M., Savchenko, I.V. (2020) Morphological genetic markers in plants. *Russian Journal of Genetics*, 56: 1406-1415.
- Clavel, J., Merceron, G., Escarguel, G. (2014) Missing data estimation in morphometrics: how much is too much? *Systematic Biology*, 63: 203-218.
- Collin, H., Fumagalli, L. (2011) Evidence for morphological and adaptive genetic divergence between lake and stream habitats in European minnows (*Phoxinus phoxinus*, Cyprinidae). *Molecular Ecology*, 20(21): 4490-4502.
- Corander, J., Marttinen, P., Sirén, J., Tang, J. (2008a) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics*, 9: 539.
- Corander, J., Sirén, J., Arjas, E. (2008b) Bayesian spatial modeling of genetic population structure. *Computational Statistics*, 23: 111-129.
- Corander, J., Waldmann, P., Sillanpää, M.J. (2003) Bayesian analysis of genetic differentiation between populations. *Genetics*, 163: 367-374.
- Cushman, S.A., McRae, B-H., McGarigal, K. (2016) *Basics of landscape ecology: an introduction to landscapes and population processes for landscape geneticists*. U:

- Balkenhol, N., Cushman, S.A., Storfer, A.T., Waits, L.P. (Urednici) Landscape genetics: Concepts, methods, applications. John Wiley & Sons. pp: 11-34.
- Deitloff, J., Petersen, J.D., Adams, D.C. (2013) Complex species interactions lead to unpredictable outcomes in *Plethodon*. *Herpetologica*, 62: 1-10.
- Diniz-Filho, J.A.F., Soares, T.N., Lima, J.S., Dobrovolski, R., Landeiro, V.L., de C Telles, M.P., Rangel, T.F., Bini, L.M. (2013) Mantel test in population genetics. *Genetics and Molecular Biology*, 36: 475-485.
- Dryden, I.L. (2021) shapes package. R Foundation for Statistical Computing, Vienna, Austria. Contributed package. Version 1.2.6.
- Dryden, I.L., Mardia, K.V. (1998) Statistical shape analysis. John Wiley & Sons.
- Dryden, I.L., Mardia, K.V. (2016) Statistical shape analysis, with applications in R (2nd ed.). John Wiley & Sons.
- Durand, E., Jay, F., Gaggiotti, O.E., François, O. (2009) Spatial inference of admixture proportions and secondary contact zones. *Molecular Biology and Evolution*, 26: 1963-1973.
- Dyer, R.J. (2014) gstudio: analyses and functions related to the spatial analysis of genetic marker data. R package, version 1.3.
- Earl, D.A., von Holdt, B.M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2): 359-361.
- Elliott, N.G., Haskard, K., Koslow, J.A. (1995) Morphometric analysis of orange roughy (*Hoplostethus atlanticus*) of the continental slope of southern Australia. *Journal of Fish Biology*, 46: 202-220.
- Escoufier, Y. (1973) Le traitement des variables vectorielles. *Biometrics*, 29: 751-760.
- Evanno, G., Regnaut, S., Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14: 2611-2620.
- Excoffier, L., Heckel, G. (2006) Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics*, 7: 745-758.
- Excoffier, L., Lischer, H.E.L. (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10: 564-567.
- Falush, D., Stephens, M., Pritchard, J.K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164: 1567-1587.
- Falush, D., Stephens, M., Pritchard, J.K. (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, 7: 574-578.
- Fontanesi, L. (2016) Metabolomics and livestock genomics: Insights into a phenotyping frontier and its applications in animal breeding. *Animal Frontiers*, 6(1): 73-79.
- François, O., Ancelet, S., Guillot, G. (2006) Bayesian clustering using hidden markov random fields in spatial population genetics. *Genetics*, 174: 805-816.

- François, O., Waits, L.P. (2016) Clustering and assignment methods in landscape genetics. U: Balkenhol, N., Cushman, S.A., Storfer, A.T., Waits, L.P. (Urednici) Landscape genetics: Concepts, methods, applications. John Wiley & Sons. pp: 114-128.
- Francuski, Lj., Djuracic, M., Ludoški, J., Milankov, V. (2013) Landscape genetics and spatial pattern of phenotypic variation of *Eristalis tenax* across Europe. Journal of Zoological Systematics and Evolutionary Research, 51: 227-238.
- Francuski, Lj., Djuracic, M., Ståhls, G., Milankov, V. (2014) Landscape genetics and wing morphometrics show a lack of structuring across island and coastal populations of the dronefly in the Mediterranean. Journal of Zoology, 292(3): 156-169.
- Francuski, Lj., Ludoški, J., Lukač, M., Dekić, R., Milankov, V. (2019) Integrative study of population structure of *Telestes dabar*, the strictly endemic cyprinid species from the Dinaric karst on the Balkan Peninsula. European Journal of Wildlife Research, 65: 66.
- Francuski, Lj., Ludoški, J., Lukač, M., Milankov, V. (2020) Fine scale population structure of hoverfly pollinator, *Eristalis arbustorum*: an integrative study. Journal of Insect Conservation, 24: 49-63.
- Francuski, Lj., Ludoški, J., Milutinović, A., Krtinić, B., Milankov, V. (2021) Comparative phylogeography and integrative taxonomy of *Ochlerotatus caspius* (Pallas, 1771) and *O. dorsalis* (Meigen, 1830). Journal of Medical Entomology, 58(1): 222-240.
- Frankham, R., Ballou, J.D., Briscoe, D.A. (2010) Introduction to Conservation Genetics. Cambridge University Press, New York.
- Gaines, W.L., Harrod, R.J., Lehmkuhl, J.F. (1999) Monitoring biodiversity: quantification and monitoring. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.
- Gasparyan, L., Mazo, I., Simonyan, V., Gasparyan, F. (2019) DNA sequencing: current state and prospects of development. Open Journal of Biophysics, 9: 169-197.
- Gojković, N., Ludoški, J., Milankov, V. (2022) The encounter of distinct *Morimus asper* (Coleoptera: Cerambycidae) phylogeographic lineages on the Balkan Peninsula: conservation implications. Journal of Insect Conservation, 26: 773-792.
- Goodall, C.R. (1991) Procrustes methods in the statistical analysis of shape. Journal of the Royal Statistical Society B, 52: 285-339.
- Goudet, J. (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. Journal of Heredity, 86: 485-486.
- Guillot, G., Mortier, F., Estoup, A. (2005) Geneland: A computer package for landscape genetics. Molecular Ecology Notes, 5: 712-715.
- Guillot, G., Rousset, F. (2013) Dismantling the Mantel tests. Methods in Ecology and Evolution, 4: 336-344.
- Gunz, P., Mitteroecker, P. (2013) Semilandmarks: A method for quantifying curves and surfaces. Hystrix, 24(1): 103-109.
- Gunz, P., Mitteroecker, P., Bookstein, F.L. (2005) Semilandmarks in three dimensions. U: Slice, D.E. (Urednik) Modern morphometrics in physical anthropology. Kluwer Press. pp: 73-98.
- Gunz, P., Mitteroecker, P., Neubauer, S., Weber, G.W., Bookstein, F.L. (2009) Principles for the virtual reconstruction of hominin crania. Journal of Human Evolution, 57(1): 48-62.

- Hall, B.K. (2003) Unlocking the black box between genotype and phenotype: cell condensations as morphogenetic (modular) units. *Biology and Philosophy*, 18: 219-247.
- Hall, L.A., Beissinger, S.R. (2014) A practical toolbox for design and analysis of landscape genetics studies. *Landscape Ecology*, 29: 1487-1504.
- Hallgrímsson, B., Percival, C.J., Green, R., Young, N.M., Mio, W., Marcucio, R. (2015) Morphometrics, 3D imaging, and craniofacial development. *Current Topics in Developmental Biology*, 115: 561-597.
- Hammer, Ø., Harper, D.A.T., Ryan, P.D. (2001) PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, 4: 9.
- Heather, J.M., Chain, B. (2016) The sequence of sequencers: the history of sequencing DNA. *Genomics*, 107(1): 1-8.
- Hedrick, P.W. (2011) *Genetics of populations*, 4th ed. Jones & Bartlett Publishers, Boston.
- Houle, D., Govindaraju, D.R., Omholt, S. (2010) Phenomics: the next challenge. *Nature Reviews Genetics*, 11: 855-866.
- <https://gomeasure3d.com/microscribe/>
- https://morphometrics.uk/MorphoJ_guide/frameset.htm?index.htm
- https://morphometrics.uk/MorphoJ_page.html
- <https://sbmorphometrics.org/>
- Jombart, T., Devillard, S., Balloux, F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11: 94-109.
- Jombart, T., Devillard, S., Dufour, A.B., Pontier, D. (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, 101: 92-103.
- Kadirvel, P., Senthilvel, S., Geethanjali, S., Sujatha, M., Varaprasad, K.S. (2015) Genetic markers, trait mapping and marker-assisted selection in plant breeding. U: Bahadur, B., Venkat Rajam, M., Sahijram, L., Krishnamurthy, K. (Urednici) *Plant biology and biotechnology: Volume II: Plant genomics*. Springer, New Delhi. pp: 65-88.
- Karp, A., Seberg, O., Buiatti, M. (1996) Molecular techniques in the assessment of botanical diversity. *Annals of Botany*, 78: 143-149.
- Kendall, D.G. (1977) The diffusion of shape. *Advances in Applied Probability*, 9: 428-430.
- Klingenberg, C.P. (2004) Integration, modules and development: molecules to morphology to evolution. U: Pigliucci, M., Preston, K. (urednici) *Phenotypic integration: studying the ecology and evolution of complex phenotypes*. Oxford University Press. pp: 213-230.
- Klingenberg, C.P. (2009) Morphometric integration and modularity in configurations of landmarks: tools for evaluating a-priori hypotheses. *Evolution & Development*, 11: 405-421.
- Klingenberg, C.P. (2010) Evolution and development of shape: Integrating quantitative approaches. *Nature Reviews Genetics*, 11(9): 623-635.
- Klingenberg, C.P. (2011) MorphoJ: An integrated software package for geometric morphometrics. *Molecular Ecology Resources*, 11: 353-357.
- Klingenberg, C.P. (2013) Visualizations in geometric morphometrics: How to read and how to make graphs showing shape changes. *Hystrix*, 24(1): 15-24.

- Klingenberg, C.P. (2015) Analyzing fluctuating asymmetry with geometric morphometrics: Concepts, methods, and applications. *Symmetry*, 7(2): 843-934.
- Klingenberg, C.P. (2019) Phenotypic plasticity, developmental instability, and robustness: the concepts and how they are connected. *Frontiers in Ecology and Evolution*, 7: 56.
- Klingenberg, C.P. (2020) Walking on Kendall's shape space: understanding shape spaces and their coordinate systems. *Evolutionary Biology*, 47: 334-352.
- Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., Mayrose, I. (2015) CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5): 1179-1191.
- Krieger, J.D. (2010) Controlling for curvature in the quantification of leaf form. U: Elewa, A.M.T. (Urednik) *Morphometrics for non-morphometricians*. Springer. pp: 27-72.
- Kumar, L.S. (1999) DNA markers in plant improvement: An overview. *Biotechnology Advances*, 17: 143-182.
- Laffont, R., Firmat, C., Alibert, P., David, B., Montuire, S., Saucedo, T. (2011) Biodiversity and evolution in the light of morphometrics: From patterns to processes. *Comptes Rendus Palevol*, 10: 133-142.
- Lawing, A.M., Polly, P.D. (2010) Geometric morphometrics: Recent applications to the study of evolution and development. *Journal of Zoology*, 280: 1-7.
- Librado, P., Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25: 1451-1452.
- Lishchenko, F., Jones, J.B. (2021) Application of shape analyses to recording structures of marine organisms for stock discrimination and taxonomic purposes. *Frontiers in Marine Science*, 8: 667183.
- Ludoški, J. (2002) Evolucionni odnosi vrsta *laticornis* grupe roda *Cheilosia* Meigen, 1822 (Diptera, Syrphidae). Magistarska teza. Prirodno-matematički fakultet, Univerzitet u Novom Sadu.
- Lukač, M. (2014) Prediona genetika vrsta *Eristalis tenax* i *E. arbustorum* (Diptera: Syrphidae) sa područja Banjalučke regije. Master rad. Prirodno-matematički fakultet Univerzitet u Banjoj Luci.
- Mabee, P.M. (2006) Integrating evolution and development: the need for bioinformatics in evo-devo. *BioScience*, 56(4): 301-309.
- MacLeod, N. (2013) PalaeoMath 101: Landmarks and semilandmarks: differences without meaning and meaning without difference. *Palaeontology Newsletter*, 82: 32-43.
- MacLeod, N. (2017) Morphometrics: History, development methods and prospects. *Zoological Systematics*, 42: 4-33.
- Manel, S., Gaggiotti, O.E., Waples, R.S. (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology and Evolution*, 20: 136-142.
- Manel, S., Schwartz, M.K., Luikart, G., Taberlet, P. (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, 18: 189-197.
- Manni, F., Guerard, E., Heyer, E. (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: How barriers can be detected by using Monmonier's algorithm. *Human Biology*, 76: 173-190.

- Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27: 209-220.
- Milankov, V., Francuski, Lj., Ludoški, J., Ståhls, G., Vujić, A. (2010) Estimating genetic and phenotypic diversity in a northern hoverfly reveals lack of heterozygosity correlated with significant fluctuating asymmetry of wing traits. *Journal of Insect Conservation*, 14: 77-88.
- Miller, M.P. (2005) Alleles in space (AIS): Computer software for the joint analysis of interindividual spatial and genetic information. *Journal of Heredity*, 96: 722-724.
- Mirzaei, S. (2021) Application of molecular markers in plant sciences; An overview. *Central Asian Journal of Plant Science Innovation*, 1: 192-200.
- Mitteroecker, P., Bookstein, F.L. (2011) Classification, linear discrimination, and the visualization of selection gradients in modern morphometrics. *Evolutionary Biology*, 38: 100-114.
- Mitteroecker, P., Gunz, P. (2009) Advances in geometric morphometrics. *Evolutionary Biology*, 36: 235-247.
- Mitteroecker, P., Huttegger, S. (2009) The concept of morphospaces in evolutionary and developmental biology: Mathematics and metaphors. *Biological Theory*, 4(1): 54-67.
- Møller, A.P., Swaddle, J.P. (1997) *Asymmetry, developmental stability and evolution*. Oxford, Oxford University Press.
- Monmonier, M.S. (1973) Maximum-difference barriers: an alternative numerical regionalization method. *Geographical Analysis*, 5: 245-261.
- Murphy, R.W., Sites, J.J.W., Buth, D.G., Haufler, C.H. (1996) *Proteins I: Isozyme electrophoresis*. U: Hillis, D.M., Moritz, C., Mable, B.K. (Urednici) *Molecular systematics*. Sinauer Associates, Sunderland. pp: 51-132.
- Nadeem, M.A., Nawaz, M.A., Shahid, M.Q., Doğan, Y., Comertpay, G., Yıldız, M., Hatipoglu, R., Ahmad, F., Alsaleh, A., Labhance, N., Ozkan, H., Chung, G., Baloch, F.S. (2018) DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnology & Biotechnological Equipment*, 32: 261-285.
- Niewoehner, W.A. (2005) A geometric morphometric analysis of late pleistocene human metacarpal 1 base shape. U: Slice, D.E. (Urednik) *Modern morphometrics in physical anthropology*. Kluwer Press. pp: 285-298.
- Padial, J.M., Miralles, A., De la Riva, I., Vences, M. (2010) The integrative future of taxonomy. *Frontiers in Zoology*, 7: 16.
- Paradis, E., Schliep, K., Potts, A., Winter, D. (2015) *pegas- population and evolutionary genetics analysis system*. R package, version 0.6.
- Perez, S.I., Bernal, V., Gonzalez, P.N. (2006) Differences between sliding semi-landmark methods in geometric morphometrics, with an application to human craniofacial and dental variation. *Journal of Anatomy*, 208: 769-784.
- Pigliucci, M. (2001) *Phenotypic plasticity: beyond nature and nurture*. The Johns Hopkins University Press.
- Primack, R.B. (2014) *Essentials of Conservation Biology*. Sinauer Associates, Sunderland.
- Pritchard, J.K., Stephens, M., Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, 155: 945-959.

- R Development Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>
- Raven, P.H. (2002) Science, sustainability, and the human prospect. *Science*, 297: 954-958.
- Rohlf, F.J. (1990) Fitting curves to outlines. U: Rohlf, F.J., Bookstein, F.L. (Urednici) Proceedings of the Michigan morphometrics workshop. University of Michigan Museum of Zoology. Ann Arbor. pp: 167-177.
- Rohlf, F.J. (2015) The tps series of software. *Hystrix*, 26(1): 9-12.
- Rohlf, F.J., Archie, J.W. (1984) A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera: Culicidae). *Systematic Zoology*, 33(3): 302-317.
- Rohlf, F.J., Corti, M. (2000) The use of two-block partial least-squares to study covariation in shape. *Systematic Biology*, 49: 740-753.
- Rohlf, F.J., Marcus, L.F. (1993) A revolution in morphometrics. *Trends in Ecology and Evolution*, 8: 129-132.
- Rohlf, F.J., Slice, D.E. (1990) Extensions of the procrustes method for the optimal superimposition of landmarks. *Systematic Biology*, 39: 40-59.
- Rousset, F. (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, 145: 1219-1228.
- Rousset, F. (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, 13: 58-62.
- Rousset, F. (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, 8: 103-106.
- Royal Society (2003) Measuring biodiversity for conservation. Policy document 11/03.
- Safner, T., Miller, M.P., McRae, B.H., Fortin, M.J., Manel, S. (2011) Comparison of Bayesian clustering and edge detection methods for inferring boundaries in landscape genetics. *International Journal of Molecular Sciences*, 12: 865-889.
- Schlager, S. (2017) Morpho and Rvcg - Shape Analysis in R. U: Zheng, G., Li, S., Szekely, G. (Urednici) *Statistical Shape and Deformation Analysis*. Academic Press. pp: 217-256.
- Schlusser, G., Wagner, G.P. (2004) *Modularity in development and evolution*. University of Chicago Press.
- Schneider, C.A., Rasband, W.S., Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9: 671-675.
- Sheets, H.D. (2014) *Integrated Morphometrics Package (IMP) 8*.
- Slatkin, M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, 139: 457-462.
- Slice, D.E. (2005) *Modern morphometrics in physical anthropology*. Kluwer Press.
- Smith, G.R. (1990) Homology in morphometrics and phylogenetics. U: Rohlf, F.J., Bookstein, F.L. (Urednici) Proceedings of the Michigan Morphometrics Workshop. University of Michigan Museum of Zoology, Ann Arbor, MI. pp: 325-338.
- Sneath, P., Sokal, R. (1973) *Numerical taxonomy: the principles and practice of numerical classification*. W.H. Freeman, San Francisco.
- Sokal, R.R., Rohlf, F.J. (1995) *Biometry: The principles and practice of statistics in biological research*. 3rd ed. W.H. Freeman and Co, New York.

- Sotola, V.A., Craig, C.A., Pfaff, P.J., Maikoetter, J.D., Martin, N.H., Bonner, T.H. (2019) Effect of preservation on fish morphology over time: Implications for morphological studies. *PLOS One*, 14(3): e0213915.
- Storfer, A., Murphy, M.A., Evans, J.S., Goldberg, C.S., Robinson, S., Spear, S.F., Dezzani, R., Delmelle, E., Vierling, L., Waits, L.P. (2007) Putting the “landscape” in landscape genetics. *Heredity*, 98: 128-142.
- Strauss, R.E. (2010) Discriminating groups of organisms. U: Elewa, A.M.T. (Urednik) *Morphometrics for non-morphometricians*. Springer. pp: 73-91.
- Strauss, R.E., Atanassov, M.N. (2006) Determining best complete subsets of specimens and characters for multivariate morphometric studies in the presence of large amounts of missing data. *Biological Journal of the Linnean Society*, 88: 309-328.
- Tatsuta, H., Takahashi, K.H, Sakamaki, Y. (2018) Geometric morphometrics in entomology: Basics and applications. *Entomological Science*, 21: 164-184.
- te Pas, M.F.W., Madsen, O., Calus, M.P.L., Smits, M.A. (2017) The importance of endophenotypes to evaluate the relationship between genotype and external phenotype. *International Journal of Molecular Sciences*, 18: 472.
- TIBCO Software Inc. (2018) Data Science Workbench, version 13. <http://tibco.com>
- Tong, Y.-J., Yang, H.-D., Jenkins Shaw, J., Yang, X.-K., Bai, M. (2021) The relationship between genus/species richness and morphological diversity among subfamilies of jewel beetles. *Insects*, 12: 24.
- Tonkin-Hill, G, Lees, J.A., Bentley, S.D., Frost, S.D.W., Corander, J. (2019) Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Research*, 47: 5539-5549.
- Valentin, A., Penin, X., Chanut, J.-P., Sevigny, J.-M., Rohlf, F.J. (2008) Arching effect on fish body shape in geometric morphometric studies. *Journal of Fish Biology*, 73: 623-638.
- Vervust, B., Van Dongen, S., Van Damme, R. (2009) The effect of preservation on lizard morphometrics- an experimental study. *Amphibia-Reptilia*, 30(3): 321-329.
- Volkova, P.A., Kasatskaya, S.A., Boiko, A.A., Shipunov, A.B. (2010) Stability of leaf form and size during specimen preparation of herbarium specimens. *Feddes Repertorium*, 121: 219-225.
- Wagner, H.H., Fortin, M.-J. (2016) Basics of spatial data analysis: linking landscape and genetic data for landscape genetic studies. U: Balkenhol, N., Cushman, S.A., Storfer, A.T., Waits, L.P. (Urednici) *Landscape genetics: Concepts, methods, applications*. John Wiley & Sons. pp: 77-98.
- Waits, L.P., Storfer, A. (2016) Basics of population genetics: quantifying neutral and adaptive genetic variation for landscape genetic studies. U: Balkenhol, N., Cushman, S.A., Storfer, A.T., Waits, L.P. (Urednici) *Landscape genetics: Concepts, methods, applications*. John Wiley & Sons. pp: 35-57.
- Wärmländer, S.K.T.S., Garvin, H., Guyomarc'h, P., Petaros, A., Sholts, S.B. (2019) Landmark typology in applied morphometrics studies: what's the point? *Anatomical Records*, 302: 1144-1153.
- Warnes, G. (2015) Genetics – population genetics. R package, version 1.3.8.1
- Warwick, A.R., Beerli, P. (2015) <http://www.peterbeerli.com/classes/images/b/ba/Geneland.pdf>

- Watson, D.F. (1992) Contouring: a guide to the analysis and display of spatial data. Pergamon Press, New York.
- Watson, D.F., Philips, G.M. (1985) A refinement of inverse distance weighted interpolation. *Geo-processing*, 2: 315-327.
- Weber, G.W., Bookstein, F.L. (2011) *Virtual anthropology: a guide to a new interdisciplinary field*. Springer.
- Webster, M., Sheets, H.D. (2010) A practical introduction to landmark-based geometric morphometrics. *Quantitative Methods in Paleobiology*, 16: 163-188.
- Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski, J.A., Tingey, S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, 18: 6531-6535.
- Wolf, J.B. (2002) The geometry of phenotypic evolution in developmental hyperspace. *The Proceedings of the National Academy of Sciences*, 99(25): 15849–15851.
- Womble, W. H. (1951) Differential systematics. *Science*, 114: 315-322.
- Wright, S. (1951) The genetical structure of populations. *Annals of Eugenics*, 15: 323-354.
- Wright, S. (1969) *Evolution and the Genetics of Populations*, Vol. II. University of Chicago Press, Chicago.
- Yazdi, F.T., Colangelo, P., Adriaens, D. (2015) Testing a long-standing hypothesis on the relation between the auditory bulla size and environmental conditions: a case study in two jird species (Muridae: *Meriones libycus* and *M. crassus*). *Mammalia*, 79: 185-200.
- Zelditch, M.L., Swiderski, D.L., Sheets, D.S. (2012) *Geometric morphometrics for biologists: A primer*. Elsevier Academic Press.