MASTERARBEIT | MASTER'S THESIS

Titel | Title

Lepidurus apus (L., 1758): A population genetic analysis through space and time

verfasst von | submitted by

Augustina Sophia Prinzessin zu Löwenstein-Wertheim-Rosenberg B.Sc.

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 831
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Zoologie
Betreut von | Supervisor:	Mag. Dr. Luise Kruckenhauser Privatdoz.
Mitbetreut von | Co-Supervisor:	Dr. Martin Schwentner

ACKNOWLEDGEMENTS

I would like to thank Martin Schwentner for introducing me to another remarkably diverse Crustacea group and for the huge opportunity to expand my knowledge in molecular research while working with a natural history collection (my favourite ones). His unwavering patience, constant support, encouragement, and mentorship, along with his expertise, were essential to this thesis. Additionally, his dedication as curator and his efforts in sampling made this research possible.

I am also incredibly grateful to Luise Kruckenhauser, whose guidance and extensive knowledge provided both direction and inspiration; without her support, I would not have had the possibility to write at the Natural History Museum.

Special thanks to Martin Kapun for his invaluable help in developing the bioinformatic pipelines and scripts for this project.

A huge thank you goes to the amazing team of the Central Research Laboratories at the NHMW, especially Alex Wanka and Rui Chen, whose expertise, trouble-shooting skills, and precious support were vital throughout my laboratory work. Many thanks to Sandra Kirchner for her time and knowledge on the MinION.

I would also like to express my gratitude to the entire Third Zoology Department, with special mention to Anna Barta for her never-ending help and Matthäus Greilhuber for his insights into large Branchiopoda and for the PhD project funding that supported my thesis.

I extend my thanks to Tobias Schernhammer and Alzbeta Devánová for their essential sampling contributions.

Finally, I am deeply indebted to my incredible family, my team-partner, and my supportive friends, whose encouragement and positivity brightened each day!

Thank you all so much!

ABSTRACT

This study investigated the population dynamics of the freshwater tadpole shrimp *Lepidurus apus* (L., 1758) (Branchiopoda, Notostraca), the only representative of this genus in Austria. Large Branchiopoda are adapted to temporary freshwater habitats via resting egg banks that persist through fluctuations and maintain populations across generations. Both spatial and temporal perspectives on the large branchiopod *L. apus* were explored: Spatially, recent genetic variation among *L. apus* populations across Austria, alongside individuals from Slovakia, was analysed to assess the extent of genetic differentiation. Temporally, the *L. apus* collection of the NHM Vienna spanning from the 1960s to the early 2000s provided a unique opportunity to investigate genetic changes over time. Population genomic methods applied ddRAD sequencing for recent material, while hyRAD was used for historic, fragmented DNA, in combination with two library preparation approaches to convert single- to double-stranded DNA, enhancing data recovery from degraded samples. Additionally, the whole genome of *L. apus* was assembled as a reference for analysing the ddRAD and hyRAD data. Results showed a clear genetic pattern in recent individuals, with Austrian and west Slovakian individuals grouping separately from those in east and south Slovakia. This pattern reflects geographic influences coupled with effective short-distance dispersal of resting eggs within the Austrian/west Slovakian group, contrasted by limited gene flow between this group and east/south Slovakia. Artefactual SNPs, detected primarily in historic libraries, were corrected to reduce their impact on genetic inference and population structure. Nonetheless, additional high missingness and low coverage made it challenging to reach clear conclusions about genetic change over time, emphasising the difficulties when working with degraded DNA in temporal studies of population dynamics.

ZUSAMMENFASSUNG

Diese Studie untersuchte die Populationsdynamik des Urzeitkrebses *Lepidurus apus* (L., 1758) (Branchiopoda, Notostraca), dem einzigen Vertreter dieser Gattung in Österreich. Groß-Branchiopoden sind an temporäre Gewässer angepasst, wobei Dauereier ihnen ermöglichen, Umweltschwankungen zu überdauern und Populationen über Generationen hinweg zu stabilisieren. Sowohl räumliche als auch zeitliche Aspekte des Groß-Branchiopoden *L. apus* wurden untersucht: Räumlich wurde die genetische Variation rezenter *L. apus* Populationen in Österreich sowie der Slowakei analysiert, um das Ausmaß genetischer Unterschiede zu bestimmen. In zeitlicher Hinsicht bot die *L. apus* Sammlung des NHM Wiens, welche Proben aus den 1960er bis frühen 2000er Jahren umfasst, eine einzigartige Gelegenheit zur Untersuchung genetischer Veränderungen im Laufe der Zeit. Populationsgenomische Methoden waren ddRAD-Sequenzierung für rezentes Material und hyRAD für historische, fragmentierte DNA ein. Die Datenerfassung wurde durch zwei Library-Vorbereitungsansätze zur Umwandlung von Einzelstrang- in Doppelstrang-DNA optimiert, um die Datengewinnung aus degradierten Proben zu verbessern. Zusätzlich wurde das gesamte Genom von *L. apus* als Referenz für die Analysen der ddRAD- und hyRAD-Daten assembliert. Die Ergebnisse zeigten ein deutliches genetisches Muster bei den rezenten Individuen, wobei die österreichischen und westslowakischen Individuen sich getrennt von denen in der Ost- und Südslowakei gruppierten. Dieses Muster spiegelt geografische Einflüsse wider, verbunden mit einer effektiven Ausbreitung der Dauereier über kurze Distanzen innerhalb der österreichisch/westslowakischen Gruppe, während der Genfluss zwischen dieser und Ost-/Südslowakei begrenzt bleibt. Artefaktische SNPs, die vorwiegend in historischen Libraries entdeckt wurden, wurden korrigiert, um ihren Einfluss auf die genetische Analyse und die Populationsstruktur zu reduzieren. Dennoch erschwerten zusätzlich hohe Fehlraten und geringe Coverage klare Schlussfolgerungen über genetische Veränderungen über die Zeit, was die Herausforderungen bei der Untersuchung degradierter DNA in einem zeitlichen Kontext zur Populationsdynamik verdeutlicht.

CONTENT

I. INTRODUCTION

I.A. Large Branchiopoda – *Lepidurus apus* (L., 1758)

Large Branchiopoda (Anostraca, Notostraca, Laevicaudata, Spinicaudata, and Cyclestherida) are uniquely adapted to life in temporary freshwater bodies, which experience seasonal cycles of filling and drying. Except Cyclestherida, which is not found in Austria, these groups are especially present in the floodplains of the Morava and Danube rivers, contributing to an outstanding diversity of large Branchiopoda in eastern Austria (Eder et al., 1997; Eder and Hödl, 2002; Eder at al., 2014).

The primary adaptation of large Branchiopods to temporary freshwater bodies is their resting eggs, which are actually resting embryos (Brendonck, 1996). These eggs are laid into the sediments during flooding events, demonstrating high drought resistance during dry periods and the ability to survive for decades (Brendonck and Meester, 2003). They are efficient agents of passive dispersal across habitats, aided by vectors such as floods, wind, and animals. However, notable genetic differentiation is observed between populations, despite their close proximity towards each other and thus high potential for gene flow (Bilton et al., 2001; De Meester et al., 2002; Vanschoenwinkel et al., 2011). This can be attributed to the Monopolisation Hypothesis (De Meester et al., 2002), which proposes that resting egg banks buffer against gene flow and newly invading genotypes. During each wet period, only a portion of the resting eggs hatch, including eggs from previous generations, resulting in an admixture of gene pools over time. This process maintains long-term genetic diversity within populations and helps prevent local extinction due to unfavourable wet seasons (De Meester et al., 2002; Evans and Dennehy, 2005; Schwentner and Richter, 2015). Thus, while resting eggs offer great potential for dispersal, they might also limit genetic exchange between populations, making it an important factor in understanding genetic diversity patterns in large Branchiopoda across Austria and neighbouring regions.

Hatching phenologies differ between Branchiopoda species; the Notostraca *Lepidurus apus* (L., 1758) hatches primarily in late spring and winter (Eder et al., 1997; Merta et al., 2016), reaching sexual maturity within few weeks (Damgaard and Olesen, 1998). Despite its widespread distribution across Europe, this tadpole shrimp is generally low in abundance. In

Austria, *L. apus* is considered near-threatened (Eder and Hödl, 2002) and occurs predominantly in the Morava floodplains. Climate change is causing irregular flood timing and duration, and decreased habitat availability and stability for *L. apus* (Löffler, 1993). Additionally, intensified land use and river regulation projects, such as the Morava restoration project ('Renaturierung Untere March-Auen'), have contributed to habitat loss (www.life-march.at).

Currently, genetic data on *L. apus* populations in Austria are limited, making it difficult to fully assess their genetic diversity. Using both historic and recent individuals from natural history collections offers a unique opportunity to examine current genetic diversity and changes over time.

I.B. Natural history collections and population genetics

Natural history collections offer a rare opportunity to explore past biodiversity, providing insights into species that are extinct, endangered, or otherwise challenging to sample today (Wall et al., 2014; Burrell et al., 2015; Derkarabetian et al., 2019; Gauthier et al., 2020; Lalueza-Fox, 2022; Fong et al., 2023).

Time-series analyses using Museomics, the large-scale genetic analysis of museum specimens, have provided valuable insights into genetic erosion, population declines, and human impacts on biodiversity over time (Burrell et al., 2015; Linck et al., 2017; Fong et al., 2023). This approach is invaluable for conservation, helping establish historical baselines of genetic diversity and track shifts in species distributions (Burrell et al., 2015; Fong et al., 2023). Although historical samples often contain highly fragmented DNA, advances in next-generation sequencing (NGS), particularly Illumina technology, enable the analysis of degraded samples by sequencing hundreds of millions of short DNA fragments simultaneously (Linck et al., 2017; Crates et al., 2019; Gauthier et al., 2020). These advancements have expanded the research potential of natural history collections, which house over 3 billion specimens globally, creating a unique spatiotemporal record of genetic diversity and biogeographic patterns (Raxworthy and Smith, 2021; Fong et al., 2023).

This study makes use of the extensive *Lepidurus apus* collection of the Natural History Museum of Vienna (NHMW), which spans from the 1960s to the early 2000s, and includes

recent individuals from 2023. Such a contiguous collection is rare and provides an exceptional opportunity to assess genetic changes over time within a single species.

I.C. RAD-Sequencing

In this study, different sequencing methods were applied to investigate the genetic diversity and population structure of *Lepidurus apus*, using both recent and historic samples. ddRAD (double-digest restriction-site associated DNA) sequencing has become a well-established method for population genetics (Peterson et al., 2012; Suchan et al., 2016; Schwentner and Lörz, 2020). Genomic DNA from individual samples is fragmented by two restriction enzymes, allowing for the sequencing and analysis of homologous fragments across individuals. This approach typically yields more than 1,000 genomic loci suitable for comparable genetic analyses.

Sequencing historic material poses several challenges as it often yields degraded and fragmented DNA, especially when treated with formalin, which not only fragments the DNA further but also introduces chemical modifications (Suchan et al., 2016). These include cytosine deamination, causing base changes such as uracil, which is read as thymine during sequencing, as well as the formation of crosslinks (protein-protein, protein-DNA, interstrand DNA crosslinks), complicating sequencing efforts (Do and Dobrovic, 2014). Much of the DNA in such samples is also single-stranded (ssDNA), requiring specialised library preparation protocols to enable a successful sequencing.

ssDNA library preparation has become essential in NGS when working with historic samples (Tin et al., 2014; Suchan et al., 2016; Linck et al., 2017; Crates et al., 2019; Gauthier et al., 2020). These methods are capable of converting ssDNA into adapter-ligated forms and efficiently recover fragments with single-strand breaks, base modifications, or short DNA fragments, which are common in historic material (Gansauge and Meyer, 2013; Kapp et al., 2021).

For this thesis, two ssDNA library preparation methods were used: *G-tailing* (Tin et al., 2014) and the *Gansauge* protocol (Gansauge and Meyer, 2013; Gansauge et al., 2020). The *G-tailing* approach ligates deoxyguanosine triphosphates (dGTPs) to ssDNA fragments, allowing a C-primer to bind and synthesise the second strand. This method is particularly useful when

working with a broad range of input DNA concentrations and requires fewer PCR cycles, making it suitable for large batches of museum samples (Tin et al., 2014; Suchan et al., 2016). In contrast, the *Gansauge* protocol implements a double-stranded adapter ligation and a biotin-streptavidin clean-up step to reduce DNA loss, making it more suitable for trace amounts of old, degraded material (Gansauge and Meyer, 2013; Kapp et al., 2021).

While converting ssDNA to dsDNA is necessary for NGS, this alone does not suffice for applying the ddRAD protocol to historic samples as the DNA fragmentation over time leads to the loss of the required restriction sites. To overcome this challenge, hybridisation restriction-site associated DNA (hyRAD) sequencing has become a favourable solution, especially when using historic material in a population genetic context (Suchan et al., 2016; Linck et al., 201; Crates et al., 2019). This method allows the capture of DNA fragments containing loci of interest using ddRAD libraries from recent samples as bait during hybridisation, enabling a subsequent sequencing of the targeted loci from historic material.

To identify homologous regions and subsequently detect genetic variation, the generated reads from both recent and historic individuals are mapped against a reference genome. As Notostracan genomes are among the smallest within the Crustaceans (~ 110 Mb) (Jeffery, 2015; Savojardo et al., 2019; Luchetti et al., 2021), Illumina short reads are sufficient for an accurate *Lepidurus apus* genome assembly (Savojardo et al., 2019; Luchetti et al., 2021). Using a species-specific genome reference further enhances the ability to detect contamination, providing a reliable framework for subsequent population genetic analyses (Savojardo et al., 2019; Blair et al., 2022a).

II. RESEARCH AIMS

The aims of the Master's Thesis are as follows:

II.A. *Lepidurus apus* through space

Understanding the patterns of gene flow and genetic differentiation among the *L. apus* populations within Austria and between Austria and Slovakia.

II.B. *Lepidurus apus* through time

Understanding the changes and potential loss of genetic diversity of *L. apus* populations in Austria over the past 60 years.

III. MATERIALS AND METHODS

III.A. Sampling, DNA-clean up and DNA extraction

For the first research aim, investigating genetic differences between *L. apus* populations within Austria and between Austria Slovakia, recent material was collected in 2013 (Austria), in 2021 (Slovakia) and 2023 (Austria and Slovakia). The Austrian individuals were obtained from various locations in the Thaya floodplain in northern Austria and in the Morava floodplain in eastern Austria, while the Slovakian individuals were collected from eastern, western, and southern regions. All individuals were fixed and stored in 80% ethanol.

To address the second research aim, detecting genetic changes and potential diversity loss in Austian *L. apus* populations over time, the historic *L. apus* collection of the NHMW was used. These samples were collected repeatedly from the Morava floodplains between the 1960s, 1990s and early 2000s, as well as from Parndorf, Burgenland. They are currently stored in 80% ethanol, but some individuals were originally fixed and stored in formalin.

The location coordinates, sampling dates, fixation, and collector of the investigated *L. apus* individuals are enlisted in the Collection Metadata Table (SI CM Tab.).

III.A.1. DNA clean-up with magnetic beads

This method is frequently used in next-generation sequencing (NGS) protocols to clean-up DNA. The volume ratio between the DNA and the beads determines the fragment lengths of

DNA that are retained: the higher the ratio of the beads' volume to DNA volume, the smaller the fragments obtained in addition to the larger ones. This is particularly useful when working with historic DNA, where a broader range of fragment sizes is often present. Due to the positively charged nanoparticles (beads), the negatively charged DNA bonds to them. The incubation time for this bonding lasted 10 minutes. By using a magnetic rack, the beads with the attached DNA were pelleted in the tube, and the supernatant, containing e.g. contaminations or a surplus of adapters, was discarded. The beads were then washed twice with fresh 80% ethanol. The ethanol was added to the tube on the magnetic rack, incubated for 30 seconds, and the supernatant was discarded. Depending on the tube's volume, 180 to 300 µl ethanol was used for each washing step. The final step of the cleanup is eluting the DNA. After removing the ethanol, the beads were allowed to dry for 5 to 10 minutes, depending on the nanoparticle volume. It is crucial to let the ethanol evaporate without completely drying out the beads. The DNA was then eluted in nucleotide-free water (NfW) or buffer, depending on the protocol, by resuspending the pellet and incubating it for 5 minutes at room temperature. The beads were pelleted in a magnetic rack, and the supernatant, containing the purified DNA, was transferred to a new tube. In the following protocols, only the beads-to-DNA volume ratio and the final elution volume are mentioned for the beads-purification steps.

III.A.2. DNA extraction

For DNA extraction, both recent and historic individuals, 1 to 10 legs were taken from the specimen and digested with 180 µl ATL Buffer from the Qiagen DNeasy Blood and Tissue Kit and 20 µl of GeneOn's Proteinase K (Tab. 1). The extracts were incubated in a thermomixer at 650 rpm for 2 to 3 hours to ensure tissue digestion. The subsequent clean-up of DNA is a necessary step after digestion to extract DNA from the tissue and remove unwanted components, such as contaminations. For the recent samples, those from the 1990s, and those older, 1x, 1.5x and 2x ratios of magnetic beads to the DNA volumes were used, respectively (Tab. 1).

Table 1: Summary of DNA extractions using SeraMag Speedbeads (magnetic beads).

Country	Collection period [year]	Number of DNA extracts per period	Tissue type & amount	ATL buffer [µl]	Protenase K [µl]	Beads' volume in ratio	Elution in Nucleotide free Water [µl]
Austria	2013-2023	7	1-3 legs	180	20	1x	32-40
Slovakia	2021-2023	23	1-3 legs	180	20	1x	32
Austria	1990-2002	54	6-7 legs	180	20	1.5x	32
Austria	1960-1970	70	10 legs	180	20	2x	32
			Incubation: 56°C for 2-3 hrs at 650 rpm				

After DNA extraction, the concentration of double-stranded DNA (dsDNA) in each extract was measured using a Quantus™ fluorometer by Promega and the QuantiFluor® ONE dsDNA System kit, following the kit's specifications. Additionally, the single-stranded DNA (ssDNA) concentration of the historic extracts was measured using a Qubit fluorometer by Thermo Fisher Scientific (Qubit™ ssDNA Assay Kit). This kit measures the concentration of all DNA; by subtracting the previously measured dsDNA, the ssDNA concentration was determined (SI LM Tab.). Overall, DNA from 154 individuals was extracted (Tab. 1).

III.B. Library preparation for double digest restriction-site associated DNA sequencing (ddRAD)

For the ddRAD sequencing, all recent *L. apus* individuals from Austria and Slovakia were selected due to the limited number of available samples and their sufficient DNA quantity (Tab. 1; SI LM Tab.). Additionally, the number was extended by eight historic DNA extracts from the 1990s and 2000s, which had an average concentration of 96 ng/µl dsDNA (SI LM Tab.). This allowed an investigation into the effectiveness of ddRAD library preparation using historic DNA (Fig. 1).

a.

Lepidurus apus through space

Country	Population	Year	Individual ID	Library ID	
Austria	Marchegg Pvt	1990	27644a	RAD02	
	Marchegg Pvt	1992	29127a	RAD02	
			29127b	RAD02	
	Zwerndorf	1994	27679a	RAD02	
			27679b	RAD02	
	Marchegg Pvt	1994	27680b	RAD02	
	Marchegg TrSe	2000	27913a	RAD02	
			27913b	RAD02	
	Austria	2013	30088	RAD15	
			30089	RAD15	
	Marchegg StW	2023	28383	RAD01	RAD10
	Marchegg Pvt	2023	28386	RAD01	RAD10
	Rabensburg	2023	28389a	RAD01	
			28389b	RAD01	WGS1
			28389c	RAD01	
Slovakia	SVK west	2021	29700a	RAD10	RAD11
			29700b	RAD11	
			29700c	RAD10	RAD11
			29700d	RAD12	
			29700e	RAD11	
	SVK east	2023	29249a	RAD13	
			29249b	RAD12	
			29701a	RAD12	
			29701b	RAD12	
			29702a	RAD13	
			29702b	RAD13	
			29702c	RAD10	RAD11
			29702d	RAD13	
			29702e	RAD13	
			29702f	RAD12	
			29702g	RAD13	
			29705a	RAD12	
			29705b	RAD13	
			29705c	RAD10	RAD11
			29705d	RAD11	
			29705e	RAD12	
	SVK south	2023	29703	RAD10	RAD11
			29704	RAD10	RAD11

b.

DNA extraction + quantification
Section III.A.2.

Whole genome sequencing
Section III.F.
WGS1
Genome assembly
Section III.G.1.a.

ddRAD
Section III.B.
RAD01 RAD02
RAD11 RAD12
RAD13 RAD15
ddRAD assembly
Section III.G.2.

Generation of biotinylated baits
Section III,D.
RAD10
hyRAD
Section III.E.

Figure 1: Overview of ddRAD workflow. Table listing each individual's collection country, location of population, sampling year, ID number and library ID for sequencing. Population abbreviations: Marchegg Pvt: Marchegg Pulverturm, Marchegg StW: Marchegg Storchen Wiese, Marchegg TrSe: Marchegg Triopssenke, SVK: Slovakia. Sampling coordinates of populations are in SI CM Tab. **b.** Method workflows for each library. Protocol names in bold, followed by corresponding section and library ID (if given).

The DNA extracts were prepared for digestion by ensuring that each extract had the same DNA amount (150 ng DNA in 25 µl). The restriction digestion reaction mix was prepared and added to the DNA extracts, resulting in a 30 µl total reaction volume (Tab. 2, Step 2). Samples were incubated in a thermocycler at 37°C, with a slow cool-down to room temperature to prevent enzyme heat-kill and maintain fragmental integrity of the library. After the restriction digestion, the digested DNA was cleaned up using magnetic beads at a 1.2x bead-to-DNA volume ratio and eluted in 21 µl NfW.

The adapters were ligated to the sticky ends of the restricted DNA fragments. Each individual received an adapter pair (MspI and EcoRI), making them unique within their pool that contained 8 to 16 samples. The adapters were added to each sample prior to the reaction mix. The ligation mix was prepared and added to the digested DNA, with a total volume of 30 µl. Ligation was performed in a thermocycler at 22°C for 1 hour, followed by heat inactivation at 65°C for 10 minutes and a slow cool-down.

The samples were pooled and cleaned using magnetic beads (1.2x volume ratio), with a final elution in 32 µl TE buffer. The size selection was performed using a BluePippin, a 2% agarose gel cassette and the marker V2, targeting a range of 280 to 550 bp. This range was based on an in-silico digestion of a published *Lepidurus packardi* genome (Blair et al., 2022a) using Geneious R10.2 (https://www.geneious.com).

In the index PCR, the adapters were completed by adding an index primer combination that makes each library pool unique, and therefore each individual (SI LM Tab.). The size-selected libraries were amplified in 4 to 6 parallel reactions to reduce PCR amplification bias, and each reaction had a total volume of 10 µl. PCR conditions included: initial denaturation at 98°C for 30 seconds, denaturation at 98°C for 10 seconds, annealing at 68°C for 75 seconds for 10-15 cycles, final extension at 70°C for 5 minutes, and cool-down to 10°C. The number of cycles was increased for some pools due to low DNA concentration after Step 6.1 (Tab. 2). However, the low DNA concentration after Step 6.1 might also have been the result of malfunctioning beads. This was discovered throughout the ddRAD protocol test phase. The following pools were amplified with an increased number of cycles to ensure a reasonable DNA concentration for sequencing: pool RAD11 with 16 cycles, and pools RAD12 and RAD13/15 with 20 cycles. The parallel reactions were pooled post-PCR and cleaned with a 1x bead volume ratio, with final

elution in 17 µl NfW. Each pool's concentration and mean fragment size were assessed using the Quantus™ fluorometer by Promega and the QuantiFluor® ONE dsDNA System kit, and the 4200 TapeStation System by Agilent. For concentration, a target of >1 ng/µl was aimed. ddRAD pools RAD02, RAD12 and RAD13/15 underwent an additional size selection (300-700 bp) due to high primer dimer content after the index PCR. All pools, except RAD10 (used only as baits for hybridisation capture, Section III.D.), were pooled at equal molarity for Illumina sequencing on the Illumina HiSeq Platform of the sequencing service provider Genohub (SI LM Tab.).

Table 2: ddRAD protocol (after Schwentner and Lörz, 2021).

Note: DNA concentrations vary across samples (SI Laboratory Metadata Table (LM Tab.)).

1. DNA preparation for digestion	
DNA	150 ng
NfW	to 25 µl
2. Restriction digestion	
2.1 Reaction master mix	**1x [µl]**
10x rCut smart buffer	3
MspI enzyme (10 000 U/ml)	0.3
EcoRI enzyme (20 000 U/ml)	0.3
NfW	1.4
DNA	25
Incubation in thermocycler: 37°C for 1-2 hrs & slow cooldown to RT (1°C/min)	
2.2 Beads clean-up: 1.2x beads & elution in 21 µl NfW	
3. Adapter ligation	
3.1 Adapter preparation	
3.1.1 Making of 10x Annealing Buffer (AB)	**1x [g]**
NaCl (500 mM)	29.22
Tris-Cl (100 mM)	15.76
3.1.2 Aliquoting of oligos from adapter stock	
100 µM stock per oligo	1 µl NfW per nmol
3.1.3 Combining complementary adapter oligos from aliquot	**100 µl adapter (40 µM)**
P1.1 adapter	40 µl
P1.2 adapter	40 µl
10x AB	10 µl
NfW	10 µl
Annealing in thermocycler: 97.5°C for 2.5 min & slow cooldown to 20°C (2.5°C/min)	
3.1.4 Diluting of adapter (40 µM) in 1x AB	
P1 (EcoRI): 10 µl stock (40 µM) in 190 µl 1x AB	200 µl of 2 µM
P2 (MspI): 79 µl stock (40 µM) in 21 µl 1x AB	100 µl of 31.6 µM
3.2 Adapter ligation	
3.2.1 Adapter with barcode per sample	**1x [µl]**
MspI adapter	3

EcoRI adapter	3
3.2.2 Reaction master mix	
10x T4 ligase buffer	3
NfW	1.25
T4 ligase	0.75
Digested DNA	19
Incubation in thermocycler: 22°C for 1 hr, 65°C for 10 min & cooldown to 10°C (2°C/90s)	
4. Pooling post adapter ligation and size selection using BluePippin	
4.1 Pooling of 8-16 samples & beads clean-up: 1.2x beads & elution in 32 µl TE buffer	
4.2 Size selection using BluePippin	
Marker V2 per sample	10 µl
Elution buffer per sample	40 µl
Agarose gel cassette	2 %
Range	280-550 bp
5. Index PCR and primer binding for Illumina	
Reaction master mix per replicate	**1x [µl]**
Primer 1 with index (10 µM)	0.5
Primer 2 with index (10 µM)	0.5
Q5 HS HiFi Master Mix	5
Library pool	4
PCR program: 98°C for 30 sec, 98°C for 10 sec, 68°C for 75 sec (12-20 cycles), 70°C for 5 min, hold at 10°C	
6. Final clean-up and equimolar pooling	
6.1 Pooling of each pool's replicates & beads clean-up: 1x beads & elution in 17 µl NfW	
6.2 Equimolar pooling for Illumina sequencing: 4nM in 80 µl	

III.C. Library preparation of historic DNA

In this Master's Thesis, two different workflows were applied to convert historic ssDNA to dsDNA: one already established in the NHMW laboratory (after Tin et al., 2014) and another newly implemented, based on the method described by Gansauge and Meyer (2013) (Fig. 3). This allowed a comparison of their handling and efficiency as they differ in synthesising the second DNA strand and the order of adapter ligation.

The first protocol (Section III.C.1.), based on Tin et al. (2014) and implemented by Barta (2023) in the NHMW laboratory, included a *G-tailing* procedure (Tab. 3, Step 2) where dGTPs were attached to ssDNA fragments. This enabled a primer consisting of six cytosine nucleotides (C-Primer) to bind and amplify the synthesise of the second DNA strand, resulting in dsDNA, to which library adapters were subsequently ligated. Three distinct approaches were used within this protocol, which are detailed in the corresponding paragraph (Section III.C.1.) (Fig. 3b).

In the *Gansauge* protocol, described by Gansauge and Meyer (2013) (Section III.C.2.), a double-stranded adapter carrying a 5' phosphate and a 3' biotin was ligated to the 3' end of the ssDNA. This reaction is possible, as part of the adapter is a splinter oligonucleotide that carries eight random nucleotides at its 3' end that complementary bind to the ssDNA fragment (Fig. 2). This workflow was extended with an additional approach using the standard Gansauge and Meyer (2013) adapters as well as ones containing a barcode.

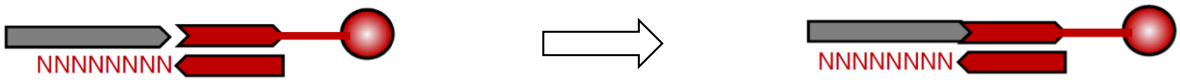


Figure 2: Schematic overview of the single-stranded library preparation protocol from Gansauge and Meyer (2013) *(Figure 1a., steps 6-9)***.** Adapter oligonucleotide (red) carrying a 5' phosphate and a 3' biotin (circle) with a splinter oligonucleotide carrying a stretch of 8 random nucleotides (N) at its 3' end and binding to ssDNA fragment (grey).

39 historic *L. apus* individuals were chosen based on their greater amount of ssDNA compared to their dsDNA concentrations (SI LM Tab.). Of this number of individuals, 24 were used for the *G-tailing* workflow (Section III.C.1.) and 15 for the *Gansauge* protocol based on Gansauge and Meyer (2013) (Section III.C.2) (Fig. 3).

a. *Lepidurus apus* through time

Country	Population	Year	Individual ID	Library prep. kit	Library ID
Austria	Marchegg	1961	27707a	FS	Gt03
			27707b	Ultra II	Gt02
			FFPE_27707b	FFPE	Gt02
			27708	Ultra II	Gt02
			FFPE_27708	FFPE	Gt02
	Marchegg	1962	26714i	FS	Gt03
	Parndorf	1962	27690a	GAB	GA1BWDH
	Parndorf	1963	27698c	GAB	GA1BWDH
	Marchegg Wgmld	1963	26712b	GA	GA1WDH
	Marchegg Pvt	1963	26718b	GA	GA1WDH
			27716a	FS	Gt03
			27716b	GA	GA1WDH
	Marchegg W	1963	27713b	GAB	GA1BWDH
	Marchegg Pvt	1964	26719a	GA	GA1WDH
			26719b	FS	Gt03
	Marchegg W	1965	26710b	FS	Gt03
	Marchegg Wgmld	1965	26711b	GA	GA1WDH
	Parndorf	1966	27693b	FS	Gt03
			FS_27686b	FS	Gt03
			GAB_27686b	GAB	GA1BWDH
	Marchegg	1967	27709	FS	Gt03
		1980	27704c	FS	Gt03
	Parndorf	1990	27646	GA	GA1WDH
			27645	GAB	GA1BWDH
	Marchegg Pvt	1990	27644a	GAB	GA1BWDH
		1992	29127c	Ultra II	Gt02
			FFPE_29127c	FFPE	Gt02
			29127cc	Ultra II	Gt02
			FFPE_29127cc	FFPE	Gt02
			29127d	Ultra II	Gt02
			FFPE_29127d	FFPE	Gt02
			29127dd	FS	Gt03
			29127e	Ultra II	Gt02
			FFPE_29127e	FFPE	Gt02
			29127ee	GA	GA1WDH
			29127f	FS	Gt03
		1995	27660a	GAB	GA1BWDH
			27660b	FS	Gt03
			27661	FS	Gt03
	Marchegg nW	1995	27677	GA	GA1WDH
	Marchegg CW	1995	27662a	FS	Gt03
			27669c	GAB	GA1BWDH
			27669d	FS	Gt03
		1997	27656b	FS	Gt03
	Baumgarten	2002	27912a	Ultra II	Gt02
			FFPE_27912a	FFPE	Gt02
			27912b	Ultra II	Gt02
			FFPE_27912b	FFPE	Gt02

b.

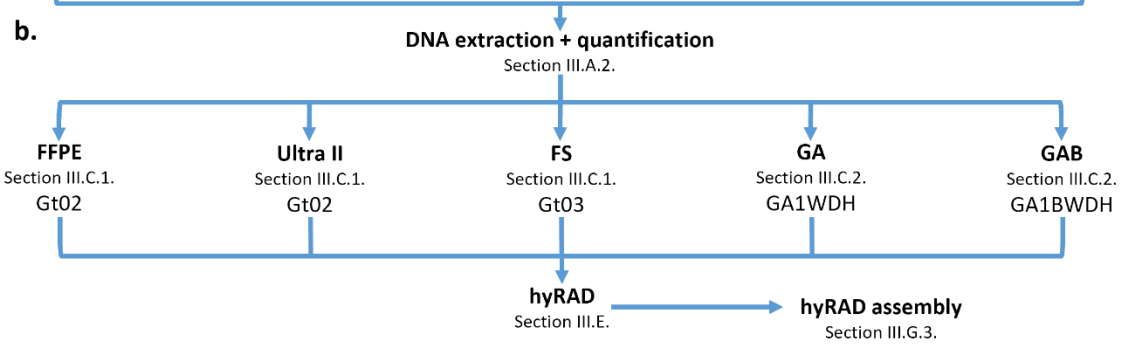


Figure 3: Overview of ssDNA library preparations and hyRAD workflow. a. Table listing each individual's collection country, location of population, sampling year, ID number, library preparation kit and library ID for sequencing. Population abbreviations: Marchegg CW: Marchegg Cyzicus-Wiese, Marchegg nW: Marchegg Wiese

nördl. Weg; Marchegg Pvt: Marchegg Pulverturm, Marchegg W: Marchegg Weide, Marchegg Wgmld: Marchegg Wegmulde. Sampling coordinates of populations are in SI CM Tab. Individual ID prefix/library preparation kit abbreviations: FFPE: FFPE+G-tailing+Ultra II (Section III.C.1., Tab. 3), Ultra II: G-tailing+Ultra II (Section III.C.1., Tab. 3), FS: G-tailing+FS (Section III.C.1., Tab. 3), GA: Gansauge (Section III.C.2., Tab. 4), GAB: Gansauge with barcoded adapter (Section III:C:2., Tab. 4). **b.** Method workflows for each library. Protocol names in bold, followed by corresponding section and library ID (if given).

III.C.1. Library preparation protocol for ssDNA using *G-tailing* after Tin et al. (2014)

Half of the DNA extracts from eight of the 24 historic individuals were treated with the NEBNext FFPE DNA Repair v2 module from New England BioLabs (here: FFPE) (Tab. 3, Step 1.), while the other half went directly into the *G-tailing* process (Tab. 3, Step 2.) (Fig. 3, library ID Gt02). The FFPE kit contains enzymes that repair single-strand breakages, crosslinks, and other DNA damages due to prior formalin fixation and paraffin embedding. This was done to assess the effectiveness of such a repair prior to library preparation. After the FFPE step, following the manufacturer's protocol with slight modifications, the repaired DNA extracts were cleaned up with a bead-to-DNA ratio of 1.5x and eluted in 20 µl NfW.

The *G-tailing* steps were performed for all 24 individuals (Tab. 3, Step 2.) (Fig. 3, library ID Gt02 & Gt03). If the initial extract volume was below 38.5 µl, it was adjusted with NfW. The DNA was then heated to 96°C for 15 minutes to degenerate double-stranded fragments into single-stranded ones. After adding the reaction mix of step 2.3 (Tab. 3), reactions were incubated for 30 minutes at 37°C, followed by heating to 70°C for 10 minutes to deactivate the transferase. The G-tailed DNA fragments were then purified with magnetic beads at a 2x ratio and eluted in 17 µl NfW.

Next, the C-primer binding and dsDNA synthesis were conducted (Tab. 3, Step 2.5). The 6C-primer, consisting of six dCTP nucleotides, hybridised to the G-tail. The Klenow fragment (Glenn et al., 2019) was used for complementary strand synthesis according to the manufacturer's instructions. The resulting dsDNA was cleaned with magnetic beads at a 2x volume ratio and eluted in various volumes of NfW, depending on the subsequent library preparation steps (Tab. 3, Step 2.6).

After the *G-tailing* process, the 16 libraries (Fig. 3, library ID Gt02), of which eight were FFPE-repaired and eight not, were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (here: Ultra II) from New England Biolabs (Tab. 3, Step 4.), whereas the remaining 16 individuals from library Gt03 were processed using NEBNext Ultra II FS DNA Library Prep Kit

for Illumina (here: FS) from New England Biolabs. The main difference between these kits is the fragmentation step performed before adapter ligation. The fragments can be cut into specific lengths depending on the incubation time, which was 15 minutes (Tab. 3, Step 3.1). This additional step was necessary because fragment measurements of the finished library Gt02 on the 4200 TapeStation System by Agilent revealed long fragments (>500 bp) (Fig. S1). Deliberate fragmentation, even of historic DNA material, is required for Illumina sequencing as fragments longer than ~ 500 bp cannot be handled.

The adapter ligation steps were the same for all libraries, but reaction volumes varied depending on the previous kit (Tab. 3, Step 3.2 or 4.2). As in the ddRAD workflow (Tab. 2, Step 3), each sample received a unique adapter within their library pool. However, instead of using adapters that ligate specifically to sticky ends produced by restriction enzymes, the adapters used here, called Y-yoke adapters, ligate to a 3' A overhang of the fragments produced by the Klenow fragment. Incubation of the adapter ligation lasted 15 minutes at 20°C, followed by bead clean-up at a 1.5x volume ratio and elution in 15 µl NfW. The index PCR was performed uniformly for all samples, assigning a unique index primer combination to each. The PCR mix and program followed the protocol (Tab. 3, Step 5). DNA concentrations were measured using the Quantus™ fluorometer by Promega and the QuantiFluor® ONE dsDNA System kit, and libraries were pooled equimolar before the final bead clean-up (0.9x volume ratio) and elution in 15 µl NfW (SI LM Tab.).

Table 3: Library preparation protocol for single-stranded DNA using G-tailing and NEBNext Ultra II FS DNA Library Prep Kit for Illumina or NEBNext Ultra II DNA Library Prep Kit for Illumina after Tin et al. (2014) and Barta (2023).

Note: DNA concentrations vary across samples (SI Laboratory Metadata Table (LM Tab.)).

1. FFPE DNA Repair	
Reaction mix	**1x [µl]**
FFPE DNA Repair buffer v2	2.33
FFPE DNA Repair mix v2	0.66
DNA	17
Incubation in thermocycler: 37°C for 15 min	
Beads clean-up: 1.5x beads & elution in 20 µl NfW	
2. G-tailing	
2.1 DNA preparation	**1x [µl]**
DNA + NfW	38.5
2.2 Heat up in thermocycler: 96°C for 15 min	
2.3 Reaction mix	**1x [µl]**

10x TdT buffer	5
$CoCl_2$ (2.5 mM)	5
Terminal transferase (20 U/µl)	0.5
dGTP (5 mM)	1
Prepared DNA	38.5
Incubation in thermocycler: 37°C for 30 min & heat up to 70°C for 10 min	
2.4 Beads clean-up: 2x beads & elution in 17 µl NfW	
2.5 C-Primer binding and dsDNA synthesis	**1x [µl]**
dNTP (10 mM each)	4.5
10x NEBuffer	3
Klenow fragment (5 U/µl)	3
6C-Primer (15 mM)	3
DNA (G-tailed)	16.5
Incubation in thermocycler: 23°C for 3h, 75°C for 20 min, hold at 10°C	
2.6 Beads clean-up: 2x beads & elution in 17-31 µl NfW	
3. Library preparation using NEBNext Ultra II FS DNA Library Prep Kit for Illumina	
3.1 DNA fragmentation	**1x [µl]**
Ultra II FS reaction buffer	3.5
Ultra II FS enzyme mix	1
DNA	13
Incubation in thermocycler: 37°C for 15 min & heated lit set to 75°C	
3.2 Adapter ligation	
3.2.1 Adapter with barcode per sample	**1x [µl]**
Y-yoke adapter (1.5 µM)	1.25
3.2.2 Reaction mix	**1x [µl]**
Ultra II ligation master mix	15
Ultra II ligation enhancer	0.5
FS reaction mixture with DNA (step 3.1)	17.5
Incubation in thermocycler: 20°C for 15 min	
3.3 Beads clean-up: 1.5x beads & elution in 15 µl NfW	
4. Library preparation using NEBNext Ultra II DNA Library Prep Kit for Illumina (alternative to 3.)	
4.1 End prep	**1x [µl]**
Ultra II end prep reaction buffer	3.5
Ultra II end prep enzyme mix	1.5
DNA	25
Incubation in thermocycler: 20°C for 30 min & heat up to 65°C for 30 min	
4.2 Adapter ligation	
4.2.1 Adapter with barcode per sample	**1x [µl]**
Y-yoke adapter (1.5 µM)	1.25
4.2.2 Reaction mix	**1x [µl]**
Ultra II ligation master mix	15
Ultra II ligation enhancer	0.5
DNA	30
Incubation in thermocycler: 20°C for 15 min	
4.3 Beads clean-up: 1.5x beads & elution in 15 µl NfW	

5. Index PCR and primer binding for Illumina	
Reaction mix	**1x [µl]**
Primer 1 with index (10 µM)	0.5
Primer 2 with index (10 µM)	0.5
Q5 HS HiFi Master Mix	5
Library pool	4
PCR program: 98°C for 30 sec, 98°C for 10 sec, 72°C for 60 sec (25 cycles), 72°C for 2 min, hold at 10°C	
6. DNA concentration measurement and equimolar pooling of libraries	
7. Beads clean-up: 0.9x beads & elution in 17 µl NfW	

III.C.2. Library preparation protocol for ssDNA after Gansauge and Meyer (2013)

The library preparation workflow followed Gansauge and Meyer (2013) with two approaches: for eight libraries the adapters designed by Gansauge and Meyer (2020) (Fig. 3, library ID GA1WDH) were used, while eight libraries received adapters with an additional barcode (Fig. 3, library ID GA1BWDH) (SI LM Tab.). The barcode approach included the DNA of individual 27686b (Parndorf, 1966), which was also part of the *G-tailing* library Gt03 (Fig. 3).

DNA extracts were adjusted to an initial volume of 30 µl with TET buffer. The first preparation step involved dephosphorylation and heat denaturation (Tab. 4, Step 2.), with reactions incubated at 37°C for 10 minutes and at 95°C for 2 minutes. During this incubation, the reaction mix for the splinted ligation of the first adapter was prepared, utilising either adapter type as described in steps 3.1 and 3.2 of the protocol (Tab. 4).

Next, the ligation products were immobilised using Streptavidin beads. Dynabeads MyOne Streptavidin C1 were resuspended, and 20 µl of beads per reaction were transferred and washed. The washing process, similar to the magnetic beads clean-up, used specific buffers (Tab. 4, Step 4.1). The beads were washed twice with B&W buffer I, resuspended, pelleted, and the supernatant discarded. Then, the beads were again resuspended in B&W buffer I and split into 100 µl aliquots. Another 100 µl of B&W Buffer I was added to the ligation reactions as excess. The diluted ligation reactions were transferred to the Streptavidin bead suspension (Tab. 4, Steps 4.2 and 4.3), and the mixture was inverted on a rotator for 20 minutes at room temperature.

Afterwards, the first of three bead washes were executed. These washings are done three times in total and in detail described here: after the inverting, the beads were pelleted in a magnetic rack and the supernatant was discarded. B&W buffer I was added, and the

supernatant was discarded after the re-pelleting of the beads. The next step was the washing with the Stringency wash buffer and incubation of the solution at 45°C for 3 minutes in a thermos mixer without shaking, followed by pelleting the beads and discarding the supernatant, then wash buffer II (Tab. 4, Step 5.) was added. While the beads stayed in wash buffer II, the reaction mix for the second-strand synthesis was prepared and added to the bead mix, after pelleting the beads and discarding the wash buffer II (Tab. 4, Step 6.). The incubation lasted for 20 minutes at 35°C in a thermos mixer with shaking at 800 rpm. Afterwards, the second bead wash was repeated as described (Tab. 4, Step 7.).

The ligation steps of the second adapter did not differ between the barcoded and non-barcoded adapters (Tab. 4, Steps 8.1 and 8.2). The beads were washed with wash buffer II, and 100 µl of the reaction mix was added to the beads, incubating for 1 hour at 22°C with shaking at 800 rpm. The final bead wash followed the same protocol as above, and the wash buffer II was discarded before eluting the DNA in 40 µl TET buffer (Tab. 4, Step 9.). The beads were resuspended and incubated at 95°C for 1 minute, then cooled to 25°C. The supernatant containing the final library was transferred to fresh tubes.

The library adapters were extended with index primers through amplification PCR, assigning a unique index primer combination to each sample. After amplification (Tab. 4, Step 11.), libraries were pooled equimolar and cleaned up with magnetic beads (0.9x bead volume to library volume) and eluted in 17 µl NfW.

Table 4: Library preparation protocol for single-stranded DNA after Gansauge and Meyer (2013).

Note: DNA concentrations vary across samples (SI Laboratory Metadata Table (LM Tab.)).

Note: The adapters and homemade buffers were prepared previously, following the 'Reagent setup' section in Gansauge et al. (2020).

1. DNA preparation	
Per DNA extract	
TET buffer	Up to 30 µl
2. Dephosphorylation and heat denaturation	
Reaction mix	**1x [µl]**
NfW	4.6
10x T4 RNA ligation buffer	8
Tween 20 (2% (vol/vol))	2
Fast AP (1 U/µl)	1
Prepared DNA	30
Incubation in thermocycler: 37°C for 10 min, 95°C for 2 min & hold at 4°C	

3. Splinted ligation of first adapter	
3.1 Reaction master mix with non-barcoded adapter	**1x [µl]**
PEG-8000 (50% (wt/vol))	32
ATP (100 mM)	0.4
TL181/TL159 (10/20 µM)	1
T4 DNA ligase (30 U/µl)	1
Dephosphorylated DNA (2. Step)	45.6
Incubation in thermocycler: 37°C for 1 h, 95°C for 2 min & hold at 10°C	
3.2 Reaction master mix with barcoded adapter	**1x [µl]**
PEG-8000 (50% (wt/vol))	32
ATP (100 mM)	0.4
TL181**_01**/TL159**_01** (10/20 µM)	1
T4 DNA ligase (30 U/µl)	1
Dephosphorylated DNA (2. Step)	45.6
Incubation in thermocycler: 37°C for 1 h, 95°C for 2 min & hold at 10°C	
4. Immobilisation of ligation products on beads	
Dynabeads MyOne Streptavidin C1 per reaction	20 µl
4.1 Washing of Streptavidin beads (2x)	
B&W buffer I	500 µl
4.2 Making 100 µl aliquots of Streptavidin beads per reaction	
Dynabeads MyOne Streptavidin C1 per reaction	1x
B&W buffer I per reaction	100 µl
B&W buffer I excess	100 µl
Split suspension into 100-µl aliquots	
4.3 Diluting of ligation reactions (3. Step) and transfer to bead suspension (4.2 Step)	**1x [µl]**
B&W buffer I	100
Ligation reaction (3. Step)	80
Bead suspension	100
Incubation on inverting rotator: room temperature for 20 min	
5. First bead wash	
5.1 Wash 1	
Ligation reaction in Streptavidin bead suspension	1x
B&W buffer I	200 µl
5.2 Wash 2	
Ligation reaction in Streptavidin bead suspension	1x
Stringency wash buffer	100 µl
Incubation in thermos mixer: 45°C for 3 min	
5.3 Wash 3	
Ligation reaction in Streptavidin bead suspension	1x
Wash buffer II	200 µl
6. Second-strand synthesis	
6.1 Reaction master mix	**1x [µl]**
NfW	38.35
10x Klenow reaction buffer	5
dNTP (25 mM each)	0.4

Tween 20 (2% (vol/vol))	1.25
CL128 (100 µM)	1
Klenow fragment (5 U/µl)	4
Ligation reaction with Streptavidin beads	1x
Incubation in thermos mixer: 35°C for 20 min & shaking at 800 rpm	
7. Second bead wash – repeating of first bead wash (5. Step)	
8. Ligation of second adapter	
8.1 Reaction master mix with non-barcoded adapter	**1x [µl]**
NfW	72.17
10x T4 DNA ligase buffer	10
PEG-4000 (50% (wt/vol))	10
Tween 20 (2% (vol/vol))	2.5
CL53/TL178 (100 µM each)	2
T4 DNA ligase (30 U/µl)	0.33
Ligation reaction with Streptavidin beads	1x
Incubation in thermos mixer: 22°C for 1 h & shaking at 800 rpm	
8.2 Reaction master mix with barcoded adapter	**1x [µl]**
NfW	72.17
10x T4 DNA ligase buffer	10
PEG-4000 (50% (wt/vol))	10
Tween 20 (2% (vol/vol))	2.5
CL53**_01**/TL178**_01** (100 µM each)	2
T4 DNA ligase (30 U/µl)	0.33
Ligation reaction with Streptavidin beads	1x
Incubation in thermos mixer: 22°C for 1 h & shaking at 800 rpm	
9. Third bead wash – repeating of first bead wash (5. Step)	
10. Elution	
Ligation reaction with Streptavidin beads (from 9. Step)	1x
TET buffer	40 µl
Incubation in thermocycler: 95°C for 1 min with heated lid & cooldown to 25°C	
11. Index PCR and amplification	
Reaction mix	**1x [µl]**
Primer 1 with index (10 µM)	5
Primer 2 with index (10 µM)	5
Q5 HS HiFi polymerase	25
Library pool	15
PCR program: 98°C for 30 sec, 98°C for 10 sec, 72°C for 60 sec (20 cycles), 72°C for 2 min, hold at 10°C	
12. DNA concentration measurement and equimolar pooling of libraries	
13. Beads clean-up: 0.9x beads & elution in 17 µl NfW	

III.D. Preparation of biotinylated baits after Lang et al. (2020)

To generate biotinylated baits, a specific ddRAD pool (Fig. 3, library ID RAD10) was created. Eight individuals were selected based on their geographic locations to ensure coverage of the genetic diversity of *L. apus* in Austria and Slovakia (SI CM Tab. and LM Tab.). The pool underwent the described ddRAD process, with a key distinction: each individual did not receive a unique adapter and index combination, but rather the same combination since the pool was not sequenced separately (SI LM Tab.). Care was taken to ensure that the combinations were unique compared to all other pools, facilitating differentiation between reads of target DNA and baits during demultiplexing after sequencing.

The bait pool was biotinylated using a biotin-carrying primer during the amplification of the first PCR (Tab. 5, Step 1.1), which also amplified the library. The bait pool was divided into eight parallel PCR reactions to minimise PCR bias. The PCR programme is detailed in Table 5. The second PCR was linear, using only the biotinylated Illumina forward primer to amplify the biotinylated strand. The reaction mix varied slightly from the first PCR, with 10 µl of biotin primer (10 µM) and 100 ng of library used (Tab. 5, Step 2.1). Between the two PCRs, and following the second, the individual reactions were pooled and cleaned using a 1.5x ratio of beads to pooled PCR reactions. The pool was eluted in 25 µl TE buffer, a medium beneficial for biotinylated DNA strands. Additionally, the concentration of the first PCR product was quantified using Quantus™ fluorometer by Promega and the QuantiFluor® ONE dsDNA System kit to determine the volume of the DNA library required for the second PCR (Tab. 5).

The measurement of the PCR product from the linear PCR was necessary to confirm whether the concentration of biotinylated baits was enough for the hybridisation. If the required 500 to 1000 ng was not reached, the linear PCR should be repeated. However, the concentration was wrongly measured as the kit did not quantify the amount of ssDNA but dsDNA; the aim of the second PCR was not only to amplify the biotinylated strands but also to prevent them from binding again. Therefore, it is uncertain how many biotinylated strands were produced and used for the hybridisation.

Table 5: Preparation protocol for biotinylated baits after Lang et al. (2020).

Note: DNA concentrations vary across samples (SI Laboratory Metadata Table (LM Tab.)).

Note: The ddRAD library RAD10 was used for making the biotinylated baits.

1. Amplification of baits via PCR	
1.1 Reaction mix	**1x**
Illumina upper biotin primer (10 µM)	5 µl
Illumina lower primer (10 µM)	5 µl
Q5 HotStart HiFi master mix	50 µl
Baits library	8 ng
NfW	Up to 100 µl
PCR program: 98°C for 30 sec, 98°C for 10 sec, 70°C for 60 sec (20 cycles), 70°C for 2 min, hold at 10°C	
1.2 Beads clean-up: 1.5x beads & elution in 25 µl TE buffer	
1.3 Concentration measurement	
2. Amplification of biotinylated strand via linear PCR	
2.1 Reaction mix	**1x**
Illumina upper biotin primer (10 µM)	10 µl
Q5 HotStart HiFi master mix	50 µl
Biotinylated baits library	200 ng
NfW	Up to 100 µl
PCR program: 98°C for 30 sec, 98°C for 10 sec, 70°C for 60 sec (20 cycles), 70°C for 2 min, hold at 10°C	
2.2 Beads clean-up: 1.5x beads & elution in 25 µl TE buffer	
2.3 Concentration measurement	

III.E. Hybridisation capture of RAD-seq loci from historic libraries (hyRAD) after Crates et al. (2019) and Lang et al. (2020)

For capturing the historic libraries which were prepared either with the *G-tailing* (Fig. 3, library ID Gt02 and Gt03) (Section III.C.1.) or *Gansauge* protocol (2013) (Fig. 3, library ID GA1WDH and GA1BWDH) (Section III.C.2.), two mixes were prepared: the library mix (LIB) and the hybridisation mix (HYB). The LIB was prepared individually for each historic library, containing 500 to 1000 ng of the equimolar pooled historic samples (converted dsDNA with unique adapters), along with various oligos for blocking non-targeted DNA or contaminants that persisted despite multiple bead clean-ups. Each LIB had a total volume of 15.6 µl (Table 6, Step 1.).

The HYB, supposedly, contained 500 to 1000 ng of the prepared biotinylated baits (Section III.D.), along with various buffer solutions and Denhardt's solution, which is a mixture of blocking agents (Tab. 6, Step 2). In a thermocycler, the double-stranded LIB was denatured by heating to 95°C for 5 minutes, followed by cooldown and holding at 65°C for 5 minutes. Simultaneously, the HYB was heated to 65°C. Maintaining the temperature at 65°C, the LIB and HYB mixes were combined into one tube, sealed with paraffin, and incubated for 48 to 72

hours, flipping twice daily (Tab. 6, Step 3). Two hybridisation runs were conducted: the first run involved hybridising the Gt02 library for 48 hours (Fig. 3, library ID Gt02), while the second run included the remaining historic libraries, which were hybridised separately, for 72 hours (Fig. 3, library IDs Gt03, GA1WDH and GA1BWDH). During the second run, the LIB-HYB-Mix of the Gt03 library evaporated and was re-eluted with 20 µl NfW after 72 hours of incubation. Concentration measurement post-incubation confirmed sufficient levels for subsequent sequencing (86 ng/µl).

Following hybridisation incubation, the historic library fragments that hybridised with biotinylated baits, were isolated using Streptavidin beads. The preparation of Streptavidin beads proceeded as follows: Dynabeads M-280 (10 mg/ml) were resuspended and washed, with the amount depending on the number of hybridisation reactions. The beads were magnetically separated, and the supernatant was removed. They were then resuspended and washed thrice in TEN buffer before eluting in 70 µl TEN (Tab. 6, Step 4).

Next, hybridised baits were attached to Streptavidin beads, and the un-hybridised library and baits were washed off. This step is crucial and sensitive, separating hybridised DNA fragments from non-target DNA (Tab. 6, Step 5). The prepared Streptavidin beads were heated to 65°C, resuspended by pipetting and combined with the LIB-HYB-Mix. This mix was incubated for 30 minutes at room temperature, flicked every 5 minutes, and then placed on a magnetic plate. The supernatant was discarded, and the beads were washed with three different washing solutions (Tab. 6, Step 5.2). The protocols of the homemade washing buffers used for the protocol are detailed in Table S1. After adding each washing solution, the bead-wash solutions were incubated for 15 minutes at 65°C, with supernatant removal and magnetic plate use between steps.

The remaining part of the protocol is after Lang et al. (2020). Following final washing and supernatant removal, beads were resuspended in 22 µl melting solution (Tab. 6, Step 5.2.4) and incubated at room temperature for 5 minutes, breaking bonds between targeted DNA fragments and biotinylated baits. During this incubation, SeraMag Speedbeads were prepared for capturing the target library (Tab. 6, Step 6.1), adding 3 µl 3M sodium acetate to the Speedbeads and acidifying them.

After the incubation of the LIB-HYB-Mix on Streptavidin beads and elution in melting solution, the released target DNA supernatant was transferred to the prepared Speedbeads. 97 µl of 96% EtOH was added and a 15-minute rotator incubation at room temperature followed. The mix was washed, air dried, eluted in 30 µl TT buffer (Tab. 6, Step 7.3), incubated for 5 minutes at room temperature and transferred to a fresh tube.

The final step was the library re-amplification of captured DNA (Tab. 6, Step 8.). Given the prior adapter ligation, this PCR used Illumina primers without indices, binding at the outer ends of the adapters, and primarily amplifying the targeted DNA from the capture. Per captured DNA pool, 5 parallel PCR reactions were prepared. Post-re-amplification, the parallel reactions were pooled, purified with beads at 1x ratio, and eluted in 17 µl NfW (Tab. 6, Step 9.).

Before the equimolar pooling for Illumina sequencing, the concentration of the captured pools was measured on Quantus™ fluorometer by Promega, with the QuantiFluor® ONE dsDNA System kit, and the fragment lengths were assessed on a 4200 TapeStation System by Agilent with the D1000 kit.

Table 6: Hybridisation capture / hyRAD protocol after Crates et al. (2019) and Lang et al. (2020).

Note: DNA concentrations vary across samples (SI Laboratory Metadata Table (LM Tab.)).

Note: The pooled libraries were made following the G-tailing protocol (Section III.C.1.) or Gansauge protocol (Section III.C.2.)).

Note: The protocol for the homemade washing buffers used for this protocol are in SI Tab. S1.

1. Library mix (LIB)	**1x [µl]**
Cot-1 DNA (1 mg/ml)	0.5
BO.1 blocking oligo (100 µM)	0.8
BO.2 blocking oligo (100 µM)	0.8
BO.3 blocking oligo (100 µM)	0.8
BO.4 blocking oligo (100 µM)	0.8
Pooled library (500-1000 ng)	11.9
Incubation in thermocycler: 95°C for 5 min & 65°C for 5 min	
2. Hybridisation mix (HYB)	**1x [µl]**
20x SSC	12
EDTA (500 mM)	0.4
SDS (10%)	0.4
50x Denhardt's solution	1.6
Biotinylated baits (500-1000 ng)	10
Prepared DNA	30

Incubation in thermocycler: 65°C for 5 min	
3. Mixing of LIB and HYB (LIB-HYB-Mix) and sealing off with paraffin	
Incubation in thermocycler: 65°C for 48-72 hrs & mixing twice each day	
4. Preparation of Streptavidin beads	
4.1 Washing of Streptavidin beads (3x)	**1x [µl]**
Dynabeads MyOne Streptavidin M-280 (10 mg/ml)	50
TEN buffer	50
4.2 Storing of Streptavidin beads	
Washed Streptavidin beads	1x
TEN buffer	70 µl
5. Attachment of hybridised baits to Streptavidin beads and washes	
5.1 Attaching hybridised baits to Streptavidin beads	
5.1.1 Preheating of Streptavidin beads: 65°C for 5-10 min	
5.1.2 Adding LIB-HYB-Mix to Streptavidin beads	
Incubation: room temperature for 30 min & flick the mix every 5 min	
5.2 Washing of LIB-HYB-Mix on Streptavidin beads	
5.2.1 Wash 1	
LIB-HYB-Mix on Streptavidin beads	1x
1x SSC/0.1% SDS	200 µl
Incubation in thermocycler: 65°C for 15 min	
5.2.2 Wash 2	
LIB-HYB-Mix on Streptavidin beads	1x
0.5x SSC/0.1% SDS	200 µl
Incubation in thermocycler: 65°C for 15 min	
5.2.3 Wash 3	
LIB-HYB-Mix on Streptavidin beads	1x
0.1x SSC/0.1% SDS	200 µl
Incubation in thermocycler: 65°C for 15 min	
5.2.4 Elution	
LIB-HYB-Mix on Streptavidin beads	1x
Melting solution	22 µl
Incubation: room temperature for 5 min	
6. Preparation of Speedbeads for capture of target library	
6.1 Washing of Speedbeads (2x)	**1x [µl]**
SeraMag Speedbeads	5
TE buffer	500
6.2 Preparing Speedbeads for capture of target library	
Washed Speedbeads	1x
NaAc (3 M)	3 µl
7. Capture of target library and washed	
7.1 Capturing target library on Speedbeads	
Pelleted supernatant of LIB-HYB-Mix on Streptavidin beads	1x
Speedbeads-NaAc mix	8 µl
96% EtOH	97 µl

Incubation on rotator: room temperature for 15 min	
7.2 Washing (2x)	
LIB-HYB-Mix on Speedbeads-NaAc mix	1x
70% EtOH	150 µl
7.3 Elution	
LIB-HYB-Mix on Speedbeads-NaAc mix	1x
TT buffer	30 µl
Incubation in thermocycler: room temperature for 5 min & transfer 28 µl in fresh tube	
8. Library re-amplification via PCR	
Reaction mix	**1x [µl]**
Illumina upper primer (10 µM)	0.5
Illumina lower primer (10 µM)	0.5
Q5 HS HiFi polymerase	5
Library pool	4
PCR program: 98°C for 30 sec, 98°C for 10 sec, 70°C for 60 sec (15 cycles), 70°C for 2 min, hold at 10°C	
9. Beads clean-up: 1x beads & elution in 17 µl NfW	
10. DNA concentration measurement and equimolar pooling of libraries for submission to an Illumina Sequencer	

III.F. Whole-genome-sequencing (WGS) of reference genome

To improve the analysis of the ddRAD and hyRAD libraries the whole genome of a recent, Austrian *L. apus* individual (Fig. 3, library ID WGS1; SI LM Tab.) was sequenced and assembled. It was planned to generate an assembly containing Illumina short reads and MinION nanopore long reads. However, the pores of both used flow cells were blocked due to unknown characteristics of the *L. apus* DNA. Consequently, the assembly is based solely on Illumina short reads, sequenced on an Illumina HiSeq Platform of the sequencing service by Genohub.

III.G. Bioinformatic pipelines

III.G.1. Genome assembly: *AutDeNovo* pipeline

The *AutDeNovo* pipeline used to assemble the *L. apus* genomes was created by Martin Kapun (© 2022 Natural History Museum Vienna). This pipeline was chosen because it has been successfully established at the NHMW and no reference genome was needed for the assembly.

To achieve the most optimal *L. apus* genome assembly for the analysis of the ddRAD and hyRAD data, four different assembly approaches were tested, each using different input data. The full *AutDeNovo* pipeline and the subsequent quality filtering steps are explained on the

basis of the first genome assembly (Section III.G.1.a.), the remaining assemblies are mentioned afterwards (Section III.G.1.b.).

III.G.1.a. *L. apus* genome assembly of recent individual from Austria

In the first steps of the pipeline, the input Illumina short reads of the recent, Austrian *L. apus* individual (Section III.F.; Fig. 3, library ID WGS1) (assembly ID: Lapus_A) underwent quality control and quality trimming using *TrimGalore* v0.6.2 (Krueger et al., 2023). The parameters used by the *AutDeNovo* pipeline were an automated adapter detection, a minimum base quality of 20 and a minimum read length of 85 bp. Subsequently, a *FASTQC* v0.11.9 (Babraham Bioinformatics - FastQC a Quality Control Tool for High Throughput Sequence Data, n.d.) output was generated to visualise the filtered and trimmed data quality. *Jellyfish* v2.3.0 (Marcais and Kingsford, 2011) and *Genomescope* v2.0 (Vurture et al., 2017) were then used to count unique k-mers and their coverage, to estimate the approximate genome size. The output was visualised in a histogram plot and a summary text file, which furthermore included the prospective ploidy level and the ratio of unique and repetitive sequences in the genome. For Illumina high-quality sequences, the de-novo assembly part of the pipeline was based on *SPAdes* v3.15.4 (Prjibelski et al., 2020) with standard parameters using the trimmed reads. The quality of the assembled genome was assessed with a combination of several analysis tools (Tab. 7).

Table 7: Quality tools used to assess the *L.* apus genome assembled with the *AutDeNovo* pipeline by Martin Kapun (NHMW).

Name of tool	Version	Authors	Description
QUAST	5.1.Orc1	Mikheenko et al. (2018)	Summarising the assembly's statistics
BLAST	2.12.0	Camacho et al. (2009)	Blasting each contig against a local copy of NCBI database to estimate off-target DNA contamination in library
BUSCO	5.2.2	Manni et al. (2021)	Database used to detect Benchmarking Universal Single-Copy Orthologs (BUSCO) from (here) Arthropods (arthropoda_odb10, *N* = 1013) in assembled scaffolds to evaluate completeness of assembly
bwa	0.7.13	Li and Durbin (2009)	Re-mapping of trimmed reads to assembled scaffolds to explore variation in read depth among different scaffolds (possible indication of off-target DNA contamination in library)
BlobTools	3.0.0	Challis et al. (2020)	Visualising quantitative analysis of assembly quality based on output of various quality tools (such as BLAST, BUSCO, bwa) by creating summary plots and tables

Post-assembly quality filtering involved excluding contigs shorter than 500 bp and removing off-target DNA contaminants, such as non-eukaryotic and human DNA, using *Python* scripts (Martin Kapun, © 2024 Natural History Museum Vienna). These scripts filtered out contigs < 500 bp and split the remaining ones according to taxon groups identified by BLAST. The eukaryotic, non-human contigs, along with those without hits in the NCBI database, were then concatenated to generate the final assembly. These filters were applied after reviewing the 'raw' assembly generated using the default *AutDeNovo* pipeline parameters.

III.G.1.b. Remaining *L. apus* genome assemblies

The second assembly (assembly ID: Lapus_Atr) was also based on these Illumina short reads, however the quality control and quality trimming of the input reads (Section III.G.1.a) were increased to a minimum base quality of 30 and a minimum read length of 100 bp (Tab. 8). This assembly was later on selected as a reference genome for the assemblies of the ddRAD and hyRAD data. For the third assembly (assembly ID: Lapus_L), *L. apus* Illumina raw reads generated by Luchetti et al. (2021) were used; the authors have published a fully annotated genome, however for comparison and evaluation reasons the genome was assembled again using the same pipeline as for the first assembly (Tab. 8). As two sets of raw reads data were available (from this Thesis and Luchetti et al., 2021), an approach to merge the data sets for better coverage and increase of completeness, was tested. Therefore, the Luchetti et al. (2021) raw reads were as strictly trimmed as the second assembly and then the trimmed forwards and reverse reads of the recent, Austrian *L. apus* individual (Fig. 3, library ID WGS1) and of Luchetti et al. (2021) were concatenated, respectively, and used as input for the *AutDeNovo* assembly pipeline (Tab. 8; assembly ID: Lapus_AL).

Table 8: Assembly parameters of the *L. apus* genomes.

Assembly ID	Data origin	Location	Collection year	Assembly parameter
Lapus_A	this Thesis (Lib. ID WGS1)	Austria, Rabensburg	2023	default *AutDeNovo* pipeline
Lapus_Atr	this Thesis (Lib. ID WGS1)	Austria, Rabensburg	2023	minimum base quality: 30; minimum read length: 100 bp; *AutDeNovo* pipeline
Lapus_L	Luchetti et al. (2021)	Austria, Marchegg	2002	default *AutDeNovo* pipeline

Assembly ID	Data origin	Location	Collection year	Assembly parameter
Lapus_AL	this Thesis + Luchetti et al. (2021)	Austria, Rabensburg + Marchegg	2023 + 2002	minimum base quality: 30; minimum read length: 100 bp; *AutDeNovo* pipeline

III.G.2. Pipeline for RAD data

III.G.2.a. Assembly of ddRAD data

The received raw reads were demultiplexed based on each individual's unique barcode (SI LM Tab.) using *cutadapt* v4.1 (Martin, 2011); PCR duplicates were removed with the *clone_filter* program of *Stacks* v2.59 (Catchen et al., 2011) and *FASTQC* v0.11.9 (Babraham Bioinformatics - FastQC a Quality Control Tool for High Throughput Sequence Data, n.d.) was used to check the quality of the raw reads. The ddRAD data was assembled by the toolkit *ipyrad* v0.9.95 (Eaton and Overcast, 2020), which has a great variety of parameters, as well as quality trimming and adapter removal, which can be adjusted to the RAD input data. For the assembly, the *ipyrad* default parameters were applied (Tab. S4), which included the minimum depth for statistical base calling of 6 reads and the minimum read length of 35 bp after adapter trimming, a reference genome-guided assembly (here: *L.* apus genome Lapus_Atr, Section III.G.1.b.), the input data type and the restriction overhang of the restriction enzymes. To decide on the thresholds for the minimum number of individuals per locus (Tab. S4, [21]) and the maximum number of SNPs per locus (Tab. S4, [22]), several test assemblies in which different combinations for [21] 80% or 90% and for [22] 20% or 10% were done (Tab. S4). For the final assembly, 90% for parameter [21] and 20% for [22] were applied. After the mapping process, *ipyrad* clustered the consensus sequences across individuals based on sequence similarity and aligns the resulting clustered sequences for the final assembly. Post assembly, SNPs which were called although they were present in less than 90% of the individuals were removed manually.

Upon later review of the ddRAD assembly, it was discovered that SNPs called in the historic individuals were likely artefacts, as they appeared only in single historic individuals. These erroneous SNPs were manually removed, but only the principal component analysis (PCA) was re-run with the corrected dataset, while the remaining genetic diversity and population structure analyses were done using the uncorrected assembly before the SNP removal.

III.G.2.b. Preparation of hyRAD data

The raw reads of the hyRAD libraries were demultiplexed as described above (Section III.G.2.a.), with great emphasis to sort out reads which had the barcoded adapters of the biotinylated baits, to ensure that only the targeted sequences were obtained. The PCR duplicates were not removed in order to retain a higher coverage. The trimming of the adapters differed based on the previously conducted ssDNA to dsDNA conversion protocol (*G-tailing* (Section III.C.1.) or *Gansauge* (Section III.C.2.)). *FASTQC* was used to examine the quality of the raw reads for downstream usage.

III.G.2.c. Assembly of hyRAD and ddRAD data

Prior to the assembly, the trimmed reads from the FFPE-repaired and non-repaired DNA libraries (Fig. 3) were concatenated (individual ID prefix 'FP', Tab. 13) to review how these combined reads compare to the original repaired and non-repaired ones, in order to assess the impact of the repair treatment on the overall quality of the data in the assembly and subsequent analyses.

The ddRAD and hyRAD sequences were mapped together against the same *L. apus* reference genome (Lapus_Atr, Section III.G.1.b) as in the *ipyrad* run, using the *bwa-aln* aligner v0.7.13 (Li & Durbin, 2006). From the *picard* tool kit v2.27.5 (http://broadinstitute.github.io/picard), the *AddOrReplaceReadGroups* was used to add the read groups to each sample's generated BAM file, and the *RealignerTargetCreator* and *IndelRealigner* of the software *GATK* (O'Connor and Van Der Auwera, 2020) were utilised to realign the mapped loci to the indels, before sorting and merging the BAM files of all mapped individuals for the variance calling. With a minimum of 30% per observed allele, *freebayes* (Garrison & Marth, 2012) produced an initial set of raw SNPs which were subsequently cleaned using *vcftools* v0.1.13 (Danecek et al., 2011), to remove indels and only keep SNPs, to retain a minimum of two alleles per SNP per individual, to exclude all sites which were not kept in the ddRAD assembly, and remove individuals which had more than 90% missingness. After filtering with *vcftools*, the VCF file was closely examined by transferring its content into an Excel sheet for manual inspection and comparison across replicates. In this sheet, the number of detected SNPs with genotypes '0/1' (heterozygous) and '1/1' (homozygous) per library was counted. When comparing the FFPE-repaired and non-repaired library replicates, unique SNPs that were present in one replicate but absent in the other and in any other hyRAD library, were identified. These SNPs, flagged

as artefacts, were removed manually from the data set, and the content was converted back into a VCF file for subsequent analyses.

The mapping script, containing the variance calling and SNP filtering, was specifically written for this Thesis, after a mapping script for NGS shot-gun sequences by Martin Kapun. A schematic overview of the script and the steps is shown in Fig. 4 and a detailed list of the utilised tools, software and command lines are in SI Table S3.

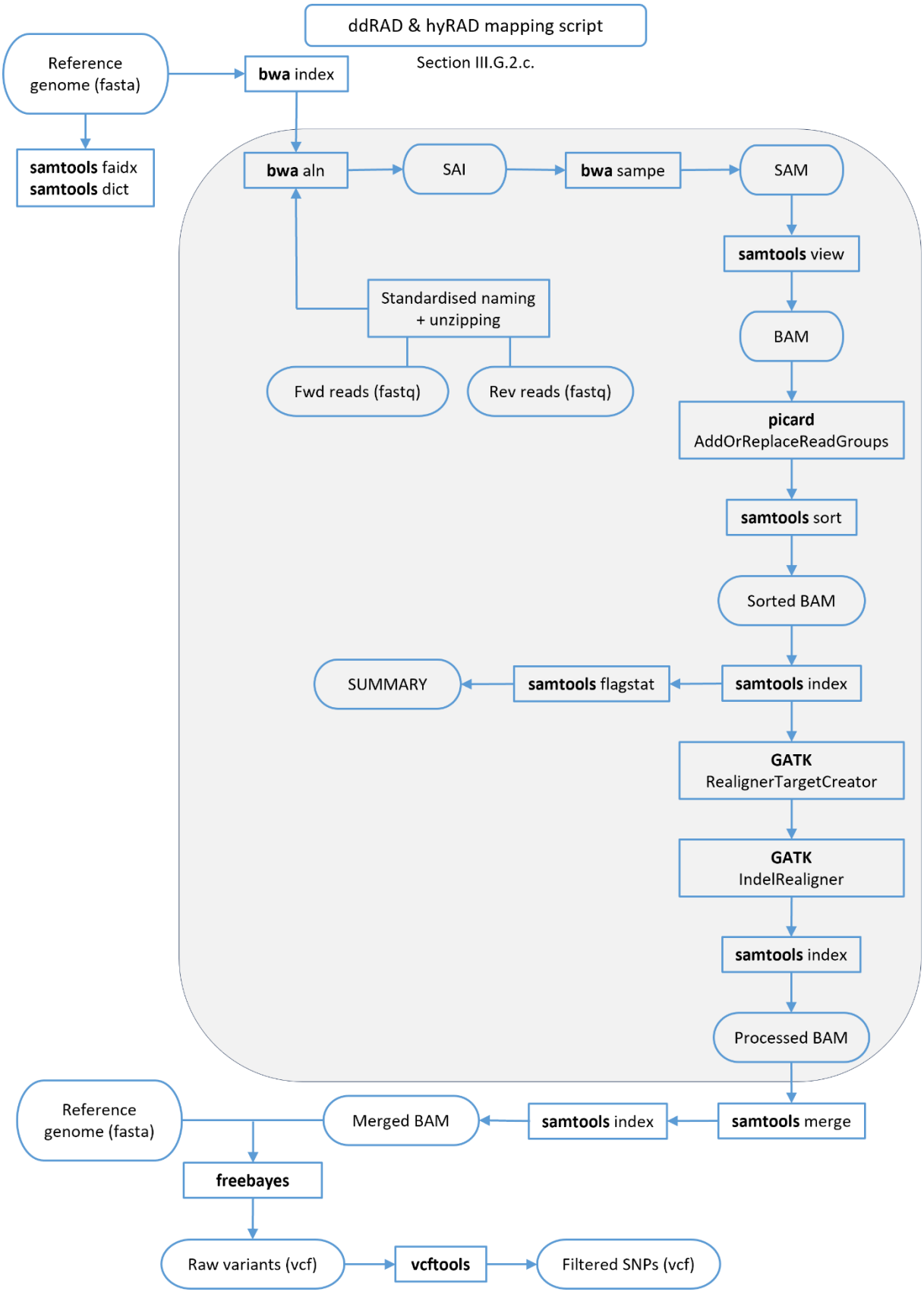


Figure 4: Schematic overview of the combined ddRAD and hyRAD data mapping script. Types of input and output data shown in round, used tools and software in square frames. All steps within the grey frame indicate the processing of each individual's sequences before merging.

III.G.3. RAD data analyses

For the purpose of a better overview and interpretation, the individuals of the ddRAD and hyRAD assemblies were not grouped in populations or time periods but kept separate. Each individual's name included its original sampling location, which corresponds to the assigned population (Fig. 1; Fig. 3), sampling year and inventory number (e.g., 'Marchegg_Pvt_2023_28386'). Therefore, the following analyses of genetic diversity and population structure were done on an individual level.

The observed (H_O) heterozygosity and population structure analyses of the ddRAD data were done based on the ddRAD assembly which included the artefactual SNPs. As they were detected upon later review, only the principal component analysis (PCA) was re-run with the corrected ddRAD assembly. For the hyRAD sub-assemblies, GC content, heterozygosity, and PCAs were calculated.

III.G.3.a. Genetic diversity analysis

The GC content was calculated via a *Python* script to inspect differences between the historic and recent individuals. This is of importance as old material, especially if they were previously fixed with formalin, have a high tendency to base modifications, which could influence the subsequent analysis of genetic diversity.

vcftools was used to calculate the SNPs based observed (H_O) heterozygosity. Due to the small number of individuals per population, H_O was calculated across the entire data set, without predefined populations. The expected (H_E) heterozygosity and the inbreeding coefficient (F_{IS}) were not calculated.

III.G.3.b. Population structure analysis

An initial assessment of the genetic relationships among *Lepidurus apus* individuals from the ddRAD assembly was done using a Neighbour-net network based on genetic distance, generated with *SplitsTree* v5.3.0 (Huson and Bryant, 2006).

Next, potential genetic structure among the Austrian and Slovakian individuals in the ddRAD data was explored through admixture analysis in *STRUCTURE* v2.3.4 (Pritchard et al., 2000). The optimal number of genetic clusters (K) was determined by testing values from 2 to 8, with a burn-in of 5,000 iterations and 50,000 generations, repeated three times for each K. The

best-fitting K was chosen based on the mean logarithmic likelihood, and admixture proportions were visualised using the R package *pophelper* (Francis, 2016).

Genetic variance was analysed in more detail by calculating principal component analysis (PCA). To compare the impact of the artefactual SNPs in the ddRAD assembly, PCA for the uncorrected and corrected assemblies were done. The PCA of the uncorrected ddRAD assembly was calculated with *GenoDive* v3.06 (© 2005-2020, Patrick Meirmans). The PCA plots were generated using the *R* package *ggplot2* v3.5.1 (Wickham et al., 2007). PCAs for the corrected ddRAD assembly and the hyRAD sub-assemblies were conducted using the *R* package *FactoMineR* v2.11 (Lê et al., 2008), with plots created using *factoextra* v1.0.7 (Kassambara & Mundt, 2016) and *ggplot2*. An additional PCA for the corrected ddRAD assembly excluded eastern and southern Slovakian individuals to focus on the clustering of Austrian and western Slovakian individuals. Similar analyses were done for the hyRAD sub-assemblies, with additional PCAs excluding historic libraries with more than 85% missing data and including only Austrian and west Slovakian libraries with less than 85% missing. Scree plots confirmed the selection of principal components, with PC1 and PC2 retained based on the percentage of explained variance.

For finer insights into the spatial genetic structure of *L. apus*, the *fineRADstructure* pipeline (Malinsky et al., 2018) was used to analyse shared ancestry among the ddRAD individuals. This analysis generated a co-ancestry matrix that summarised the nearest neighbour haplotype relationships. The co-ancestry matrix was visualised as a heatmap, accompanied by a haplotype relationship tree, providing detailed information on shared ancestry and genetic connections between individuals.

IV. RESULTS

IV.A. *L. apus* genome assembly

Four different *L. apus* genomes were assembled of which the most optimal, assembly Lapus_Atr, was chosen as a reference genome for the subsequent analysis of the ddRAD and hyRAD data (Tab. S5). As the genomes were solely assembled to generate reference for the mapping of the RAD data, no further genomic analyses were done. The outcomes of the 'raw' assemblies, prior to the filtering steps, and of the other three assemblies were checked with *QUAST* and *BUSCO* (Tab. S5); however, the results presented and discussed here are only of the final selected *L. apus* genome.

155,563,732 raw sequences were generated from the recent, Austrian *L. apus* individual (Section III.F.; Fig. 3, library ID WGS1) for the Lapus_Atr assembly. The k-mer analysis was 31 bp, estimating a genome size of 83.8 Mb with an approximate genome coverage of 187 ×, and heterozygosity of 0.05% (Fig. S5). The Lapus_Atr genome was chosen as the assembly length (84.3 Mb) matched the estimated genome size best while consisting of the smallest number of contigs (16,810) and having the greatest N50 value (63.1 kb). The GC content was 42.06% and based on the arthropods' gene set on *BUSCO* (*N* = 1030 orthologs), the result for the assembly indicated a high degree of completeness (97.3%) (Tab. S5).

IV.B. ddRAD data

IV.B.1. ddRAD assembly

Overall, 87.69 million raw reads, of which 80% were from recent *L. apus* individuals (Austria: 35%, Slovakia:65%), and 20% from historic (Tab. 9), were generated from 35 ddRAD libraries. For three libraries, no sequences were generated. This might be due to a laboratory error as they were ligated with the same MspI adapter (SI LM Tab.). After the first quality check via *FASTQC*, one historic and one recent individual were removed additionally for the following *ipyrad* assembly (Tab. 9). In total, 33 ddRAD samples, of which seven were historic, with 87,687,136 reads went into the assembly pipeline of *ipyrad*.

Throughout the assembly, additionally 14 ddRAD individuals, predominantly from Slovakia and historic ones from Austria, were excluded because *ipyrad* was unable to properly cluster

them or due to low read quality (Tab. 9). Therefore, the final assembly contained four historic, Austrian individuals, five recent ones from Austria and ten from Slovakia, with 47,986,545 mapped reads. A total of 5,301 loci, 4,065 monomorphic and 11,713 SNPs, were detected, and after the first round of manual SNP removal, the assembly had 1041 polymorphic loci with 5,163 SNPs (Tab. 9).

Due to the late review and discovery of artefactual SNPs in single historic individuals, 289 SNPs were additionally removed. As mentioned, the observed (H_O) heterozygosity and population structure analyses were done based on the uncorrected ddRAD assembly (5,163 SNPs), except the PCA which was additionally re-run with the corrected 4,874 SNPs.

IV.B.2 Genetic diversity

The GC content of each individual's mapped reads revealed that the historical samples had similar GC contents as the recent ones. This suggests that the historical DNA, despite its age, had no extensive post-mortem base mutations that could have influenced the results (Tab. 10). What is noticeable, however, is that the two recent individuals from Marchegg (Tab. 10, 28383 and 28386) have the lowest GC content with < 41.9% (Tab. 10).

The observed heterozygosity (H_O) values were calculated based on the un-corrected ddRAD assembly with 5,163 SNPS. The H_O values were overall low, ranging from 0.0219 (Tab. 10, 28389c) to 0.178 (Tab. 10, 29700d). The highest H_O value was observed for the west Slovakian individual 29700d from the year 2021. Overall, the Slovakian individuals had an average H_O of 0.063, considerably higher than the average H_O of 0.036 of the recent ones from Austria. The historic samples had a similar low H_O of 0.037 (Tab. 10).

Tab. 9: Summary of processed ddRAD sequences after demultiplexing and assembly.

Note: Complete population names are given in the figure text of Fig. 1.

Note: 'x' indicates a positive FASTQC check (for subsequent assembly) and/or a successful clustering by ipyrad *of each individual's sequences*.

Note: Number of loci and SNPs post filtering of the uncorrected ddRAD assembly.

** In regard to polymorphic and monomorphic loci.*

*** In regard to polymorphic loci only.*

Country	Population	Year	Indv. ID	Library ID	Nbr. raw reads	FASTQC check	Included by ipyrad	Nbr. reads assembled	Nbr. of loci (total 5301) *	Nbr. of SNPs (total 11713)	Nbr. of loci post filter (total 1041) **	Nbr. of SNPs post filter (total 5163)
Austria	Marchegg Pvt	1990	27644a	RAD02	7	-	-	-	-	-	-	-
	Marchegg Pvt	1992	29127a	RAD02	7306071	x	x	6455964	5263	9716	1032	5121
			29127b	RAD02	4013081	x	x	3443172	5273	9560	1032	5120
	Zwerndorf	1994	27679a	RAD02	114834	x	-	-	-	-	-	-
			27679b	RAD02	756342	x	-	-	-	-	-	-
	Marchegg Pvt	1994	27680b	RAD02	63306	x	-	-	-	-	-	-
	Marchegg TrSe	2000	27913a	RAD02	2400241	x	x	2087007	5218	8692	995	4962
			27913b	RAD02	2196830	x	x	1924615	5199	8337	973	5010
	Austria	2013	30088	RAD15	142822	x	-	-	-	-	-	-
			30089	RAD15	3450200	x	x	2598741	5275	10157	1031	5119
	Marchegg StW	2023	28383	RAD01	6312453	x	x	5320603	5211	9226	990	5022
	Marchegg Pvt	2023	28386	RAD01	4995532	x	x	3501738	5222	8912	983	4997
	Rabensburg	2023	28389a	RAD01	376	-	-	-	-	-	-	-
			28389b	RAD01	6072307	x	x	4717841	5210	9974	1016	5031
			28389c	RAD01	3985945	x	x	2247358	5280	10002	1032	5113
Slovakia	SVK west	2021	29700a	RAD11	698585	x	-	-	-	-	-	-
			29700b	RAD11	2278075	x	x	420655	4927	7712	932	4725
			29700c	RAD11	1238149	x	x	733186	5272	8934	1031	5082
			29700d	RAD12	1478363	x	x	537979	5165	7967	1005	4961

Country	Population	Year	Indv. ID	Library ID	Nbr. raw reads	FASTQC check	Included by ipyrad	Nbr. reads assembled	Nbr. of loci (total 5301) *	Nbr. of SNPs (total 11713)	Nbr. of loci post filter (total 1041) **	Nbr. of SNPs post filter (total 5163)
			29700e	RAD11	0	-	-	-	-	-	-	-
	SVK east	2023	29249a	RAD13	0	-	-	-	-	-	-	-
			29249b	RAD12	1	-	-	-	-	-	-	-
			29701a	RAD12	629750	x	-	-	-	-	-	-
			29701b	RAD12	4408839	x	x	3554721	5263	8705	1029	5119
			29702a	RAD13	644480	x	-	-	-	-	-	-
			29702b	RAD13	2335806	x	-	-	-	-	-	-
			29702c	RAD11	1496422	x	x	714860	5270	8890	1025	5086
			29702d	RAD13	932913	x	-	-	-	-	-	-
			29702e	RAD13	3234185	x	-	-	-	-	-	-
			29702f	RAD12	2313666	x	x	480708	5206	8020	1012	4990
			29702g	RAD13	1667739	x	-	-	-	-	-	-
			29705a	RAD12	562805	x	-	-	-	-	-	-
			29705b	RAD13	4584564	x	-	-	-	-	-	-
			29705c	RAD11	499164	x	-	-	-	-	-	-
			29705d	RAD11	1371490	x	x	180760	4263	6085	729	3860
			29705e	RAD12	4832175	x	x	319731	4825	7002	901	4667
	SVK south	2023	29703	RAD11	585630	x	x	179945	4679	6809	867	4455
			29704	RAD11	10084372	x	x	8566961	5260	10090	1024	5058

Table 10: Summary of SNP based statistical results to represent genetic diversity of the *L. apus* individuals in the uncorrected ddRAD assembly (1,563 SNPs).

Note: Complete population names are given in the figure text of Fig. 1.

Country	Population	Year	Indv. ID	GC content (%)	Obst. Het (H_O)
Austria	Marchegg Pvt	1992	29127a	42.25	0.0475
			29127b	42.34	0.0447
	Marchegg TrSe	2000	27913a	42.19	0.0268
			27913b	42.13	0.0285
	Austria	2013	30089	42.21	0.0276
	Marchegg StW	2023	28383	41.86	0.0472
	Marchegg Pvt	2023	28386	41.87	0.0327
	Rabensburg	2023	28389b	42.11	0.0495
			28389c	42.22	0.0219
Slovakia	SVK west	2021	29700b	42.26	0.0572
			29700c	42.17	0.0379
			29700d	42.03	0.1770
	SVK east	2023	29701b	42.05	0.0477
			29702c	42.20	0.0470
			29702f	42.11	0.0584
			29705d	42.11	0.0465
			29705e	41.98	0.0549
	SVK south	2023	29703	41.97	0.0524
			29704	42.34	0.0569

IV.B.3. Population structure

To investigate the population structure of *Lepidurus apus*, multiple methods were employed, including Neighbour-net split network, *STRUCTURE* analysis, Principal Component Analysis (PCA), and *fineRADstructure*. The PCA was done for the corrected (4,874 SNPs), whereas the remaining population analyses were done with the uncorrected ddRAD assembly (5,163 SNPs).

Each of these analyses revealed a similar clustering pattern: individuals from eastern and southern Slovakia were clearly separated from those in Austria and western Slovakia, with the latter forming a close cluster (Fig. 8).

The Neighbour-net network provided an initial visualisation of the described genetic pattern (Fig. 5), reflecting a strong geographic structure between the Austrian/west Slovakian and east/south Slovakian groups.

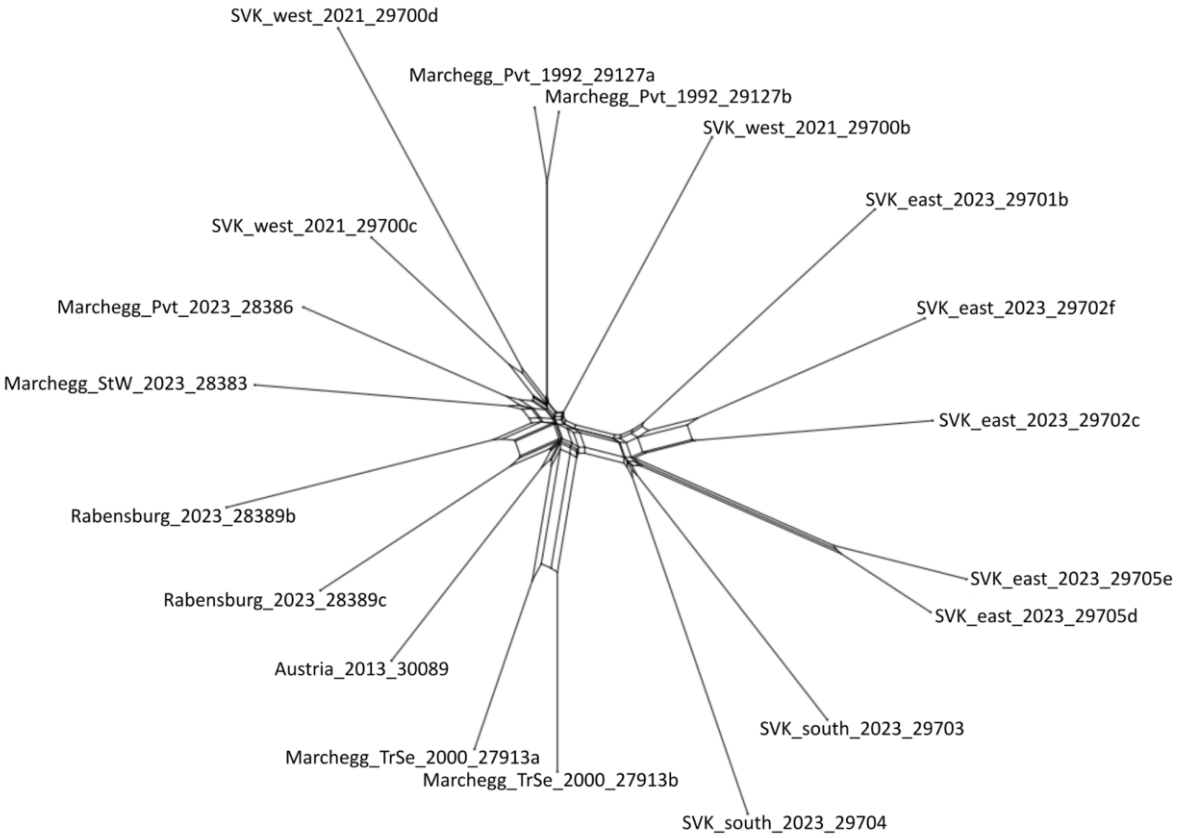


Figure 5: Midpoint-rooted Neighbour-net split network showing the genetic relationships among *Lepidurus apus* individuals from the uncorrected ddRAD assembly (5,163 SNPs). Two major clusters are observed, with the Austrian and western Slovakian individuals forming one group, distinct from the remaining Slovakian individuals.

For the admixture analysis (done with the uncorrected ddRAD assembly), it was decided to present the results of the *STRUCTURE* bar plot for K = 5, as it had the highest mean log likelihood, indicating it was the best fit for explaining the genetic structure (Fig. 6f). All individuals shared a common genetic background (Fig. 6e, purple). However, individuals from eastern and southern Slovakia were distinguishable from the remaining *L. apus* individuals, displaying a distinct genetic composition in addition to their shared ancestry (Fig. 6d, red). Overall, this genetic pattern was consistent across the plots for the remaining K values (K = 2, 3, 4, 6, 7, and 8; Fig. S3).

In the PCA analysis of the uncorrected ddRAD assembly (5,163 SNPs), the recent individuals from Marchegg clustered closely with those from western Slovakia. Further on, the PCA plot revealed a clear separation between the historic Austrian individuals and the others (Fig. 6b). The two Marchegg TrSe individuals from 2000 were positioned furthest from the main Austrian cluster on one side, while the Marchegg Pvt (1992) individuals were located on the opposite side. As a result, instead of the two major groupings as observed in the Neighbour-

net and STRUCTURE analysis, the ddRAD individuals were grouped into four distinct clusters (Fig. 6b).

The PCA of the corrected ddRAD assembly (4,874 SNPs) showed overall the same pattern, the separate clustering of east/south Slovakian individuals from Austria/west Slovakia. But there were some minor differences regarding the historic individuals: After recalculating the PCA, the individuals from Marchegg Pvt (1992) clustered more closely with the recent Austrian and western Slovakian individuals, especially with 28383 (Marchegg StW, 2023) (Fig. 6c). When removing the eastern/southern Slovakian individuals, the historic individuals from Marchegg (29127a/b and 27913a/b) were positioned again towards the outer region of the Austrian/west Slovakian cluster, the individuals 29127a/b from Marchegg Pvt (1992) with the west Slovakian and 27913a/b from Marchegg TrSe (2000) together with the Austrian individual from 2013. Noticeable, the individuals from Rabensburg (2023) were clustered furthest out (Fig. 6d).

The results from *fineRADstructure* further support the patterns observed in the other analyses, revealing two well-supported clusters corresponding to the two main geographical groups. Within the eastern/southern Slovakian group, the southern individuals exhibited greater shared recent coancestry with a subset of the eastern Slovakians (Fig. 7). Notably, the two eastern individuals 29705d and 29705e, collected from the same location in 2023, showed the most pronounced coancestry not just within their group but also compared to all *L. apus* individuals in the ddRAD assembly. Similarly, a noticeable but lower degree of coancestry was observed within the Marchegg populations from 1992 and 2000, with these individuals again separating from the rest as seen in the PCA of the uncorrected ddRAD assembly (Fig. 7b).

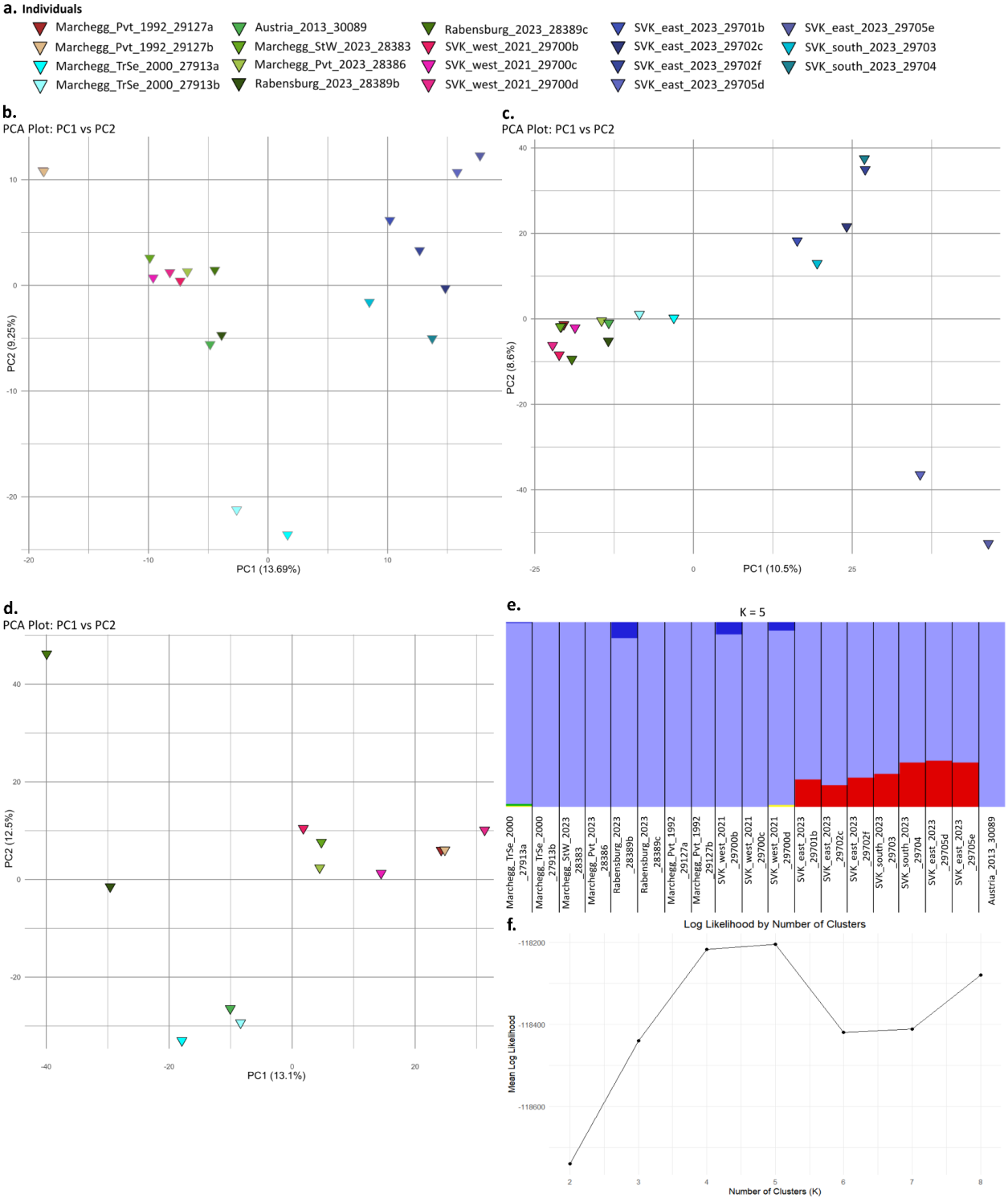


Figure 6: Principal component analysis (PCA) and STRUCTURE results visualising population structure of *L. apus* individuals based on ddRAD SNPs. a. Legend of ddRAD individuals in PCA plots (b-d). **b.** PCA of uncorrected ddRAD assembly (5,163 SNPs); individuals 29127a/b are clustered on top of each other. **c.** PCA of corrected ddRAD assembly (4,874 SNPs), excluding artefactual SNPs of historic individuals; individuals 29127a/b are clustered on top of each other. **d.** PCA for corrected ddRAD assembly, excluding artefactual SNPs of historic individuals and eastern/southern Slovakian individuals. **e**. Bar plot of admixture proportions of the 5,163 ddRAD-seq SNPS (uncorrected ddRAD assembly) for K = 5, each bar represents an individual, different colours correspond to inferred genetic clusters. **f.** Mean log likelihood values for K (number of clusters) ranging from two to eight, with three replicates per K, highest mean log likelihood was observed at K = 5. See Fig. S3 for PCA scree plots.

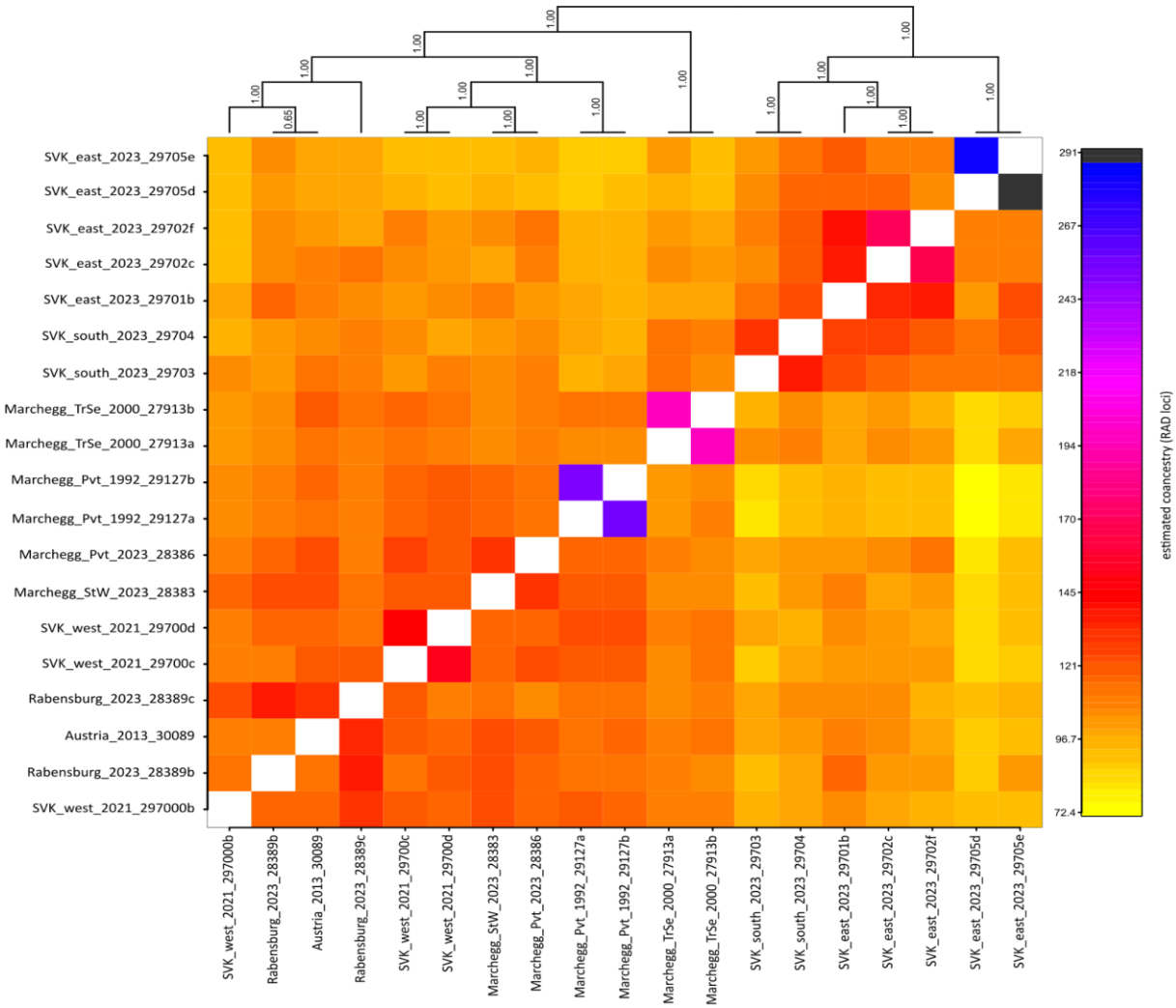


Figure 7: Coancestry matrix from *fineRADstructure* based on ddRAD loci from uncorrected *L. apus* assembly. Pairwise coefficients of coancestry are coloured from low (yellow) to high (black). Dendrogram represents a clustering of the individuals based on a pairwise matrix of coancestry coefficients.

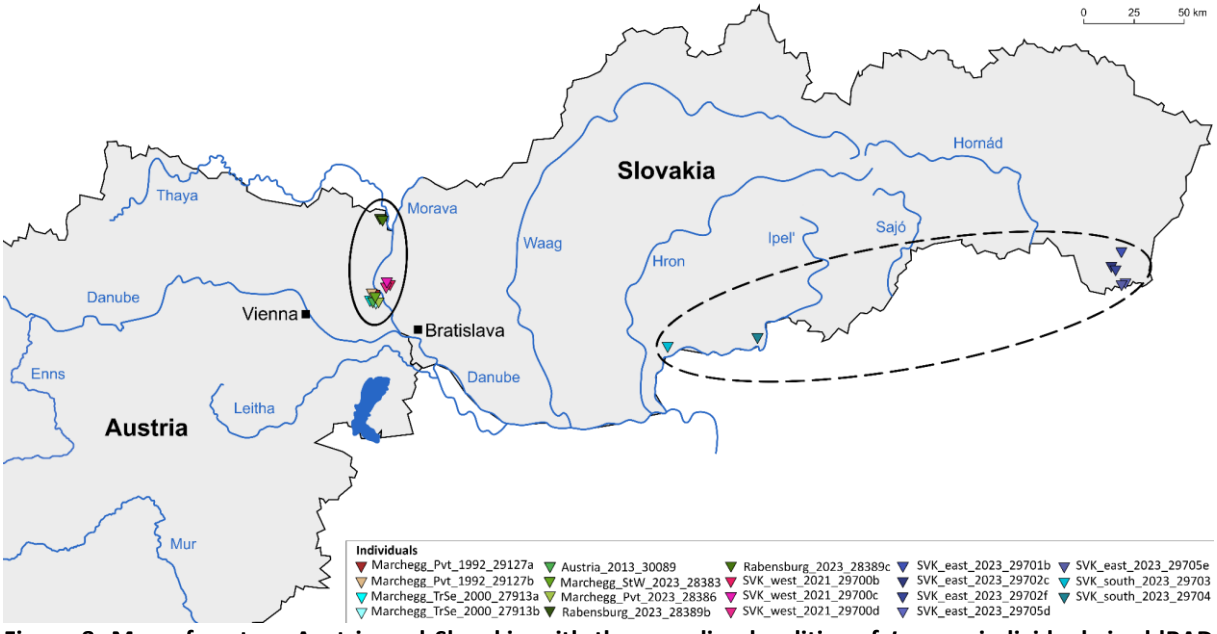


Figure 8: Map of eastern Austria and Slovakia with the sampling localities of *L. apus* individuals in ddRAD

assembly, displaying the major division between Austrian/west Slovakian cluster (solid circle) and east/south Slovakian cluster (dashed circle). Differently coloured triangles correspond to individuals.

IV.C. hyRAD data

48 hyRAD libraries from 39 historic *L. apus* individuals were sequenced, but as for five individuals the same barcode for different libraries was used, these were not demultiplexed (27698c, 27713b, 26710b, 27646, FFPE_29127cc; Tab. 11; SI LM Tab.). Therefore, raw reads from 43 libraries of 35 individuals were obtained (total 30,333,073; on average 7,054,420). Following quality trimming and checks, 18 individuals (21 libraries) were excluded, resulting in 22 hyRAD libraries from 17 individuals and 3,178,428 reads (on average: 144,474) for the assembly. Of those, eight individuals were from the 1960s, seven from the 1990s, and two from 2002 (Tab. 11).

As different numbers of DNA libraries were used per library preparation method, the average number of reads is presented below for comparison.

The *G-tailing* protocol (Tin et al., 2014) yielded 973,960 raw reads on average, while the *Gansauge* protocol (Gansauge and Meyer, 2013) produced 85,712 reads on average. After trimming and quality filtering, 16 *G-tailing* and five *Gansauge* libraries were excluded, leaving an average of 203,934 reads from *G-tailing* and 40,419 from *Gansauge* libraries (Tab. 11). Overall, more high-quality reads were obtained on average using the *G-tailing* method.

As mentioned, reads from individual FFPE_29127cc were not included due to barcode duplication. Therefore, for the FFPE-repaired libraries, 1,328,568 raw reads on average were generated, and on average 866,036 reads for the non-repaired libraries. After trimming, two libraries from each group were excluded, leaving on average 150,896 reads from the repaired DNA libraries and 256,972 from the non-repaired, suggesting the FFPE repair process was less effective than anticipated (Tab. 11).

From the *Gansauge* protocol, one library with non-barcoded adapters (library ID: GA1WDH) and two with barcoded adapters (library ID: GA1BWDH) were excluded due to barcode duplication. The libraries of GA1WDH yielded 135,664 raw reads on average, while the barcoded libraries (GA1BWDH) produced an average of 27,436 raw reads. After quality trimming, and further excluding four libraries from GA1BWDH and one from GA1WDH, 12,195 and 49,827 on average, respectively, were retained for assembly (Tab. 11).

Tab. 11: Summary of processed hyRAD sequences after demultiplexing and quality trimming.

Note: Complete population names and individual ID prefixes are given in the figure text of Fig. 3.

Library preparation kits: FFPE: FFPE+G-tailing+Ultra II (Section III.C.1., Tab. 3), Ultra II: G-tailing+Ultra II (Section III.C.1., Tab. 3), FS: G-tailing+FS (Section III.C.1., Tab. 3), GA: Gansauge (Section III.C.2., Tab. 4), GAB: Gansauge with barcoded adapter (Section III.C.2., Tab. 4).

Note: 'x' indicates a positive FASTQC check of library and selection for subsequent assembly.

Country	Population	Year	Indv. ID	Library prep. kit	Library ID	Nbr. raw reads	Nbr. reads post trimming	FAST QC check	Library used in analysis
Austria	Marchegg	1961	27707a	FS	Gt03	4411459	3712	-	-
			27707b	Ultra II	Gt02	127461	978	-	-
			FFPE_27707b	FFPE	Gt02	20361	6684	x	-
			27708	Ultra II	Gt02	173853	3527	x	-
			FFPE_27708	FFPE	Gt02	85158	2741	-	-
	Marchegg	1962	26714i	FS	Gt03	819	85	-	-
	Parndorf	1962	27690a	GAB	GA1BWDH	3520	225	-	-
	Parndorf	1963	27698c	GAB	GA1BWDH	-	-	-	-
	Marchegg Wgmld	1963	26712b	GA	GA1WDH	26686	15231	x	x
	Marchegg Pvt	1963	26718b	GA	GA1WDH	334698	163117	x	x
			27716a	FS	Gt03	77209	2413	-	-
			27716b	GA	GA1WDH	292697	27385	-	-
	Marchegg W	1963	27713b	GAB	GA1BWDH	-	-	-	-
	Marchegg Pvt	1964	26719a	GA	GA1WDH	89264	18156	x	x
			26719b	FS	Gt03	2173385	3548	-	-
	Marchegg W	1965	26710b	FS	Gt03	-	-	-	-
	Marchegg Wgmld	1965	26711b	GA	GA1WDH	58864	21876	x	x
	Parndorf	1966	27693b	FS	Gt03	32087	881	-	-
			FS_27686b	FS	Gt03	465353	25415	-	-
			GAB_27686b	GAB	GA1BWDH	110650	13567	x	x
	Marchegg	1967	27709	FS	Gt03	41085	893	-	-
		1980	27704c	FS	Gt03	230	80	-	-

Country	Population	Year	Indv. ID	Library prep. kit	Library ID	Nbr. raw reads	Nbr. reads post trimming	FAST QC check	Library used in analysis
	Parndorf	1990	27646	GA	GA1WDH	-	-	-	-
			27645	GAB	GA1BWDH	885	165	-	-
	Marchegg Pvt	1990	27644a	GAB	GA1BWDH	45503	10822	x	-
		1992	29127c	Ultra II	Gt02	556603	270053	x	x
			FFPE_29127c	FFPE	Gt02	434456	151071	x	x
			29127cc	Ultra II	Gt02	122098	18624	x	x
			FFPE_29127cc	FFPE	Gt02	-	-	-	-
			29127d	Ultra II	Gt02	5180799	329233	x	x
			FFPE_29127d	FFPE	Gt02	3422024	290189	x	x
			29127dd	FS	Gt03	1764907	29093	-	-
			29127e	Ultra II	Gt02	510446	355399	x	x
			FFPE_29127e	FFPE	Gt02	44903	5353	x	x
			29127ee	GA	GA1WDH	63347	58581	x	x
			29127f	FS	Gt03	73236	5901	x	x
		1995	27660a	GAB	GA1BWDH	2912	19	-	-
			27660b	FS	Gt03	54790	168	-	-
			27661	FS	Gt03	43995	32	-	-
	Marchegg nW	1995	27677	GA	GA1WDH	84092	22002	x	-
	Marchegg CW	1995	27662a	FS	Gt03	2038439	3940	-	-
			27669c	GAB	GA1BWDH	1143	427	-	-
			27669d	FS	Gt03	23152	1542	-	-
		1997	27656b	FS	Gt03	861580	4987	-	-
	Baumgarten	2002	27912a	Ultra II	Gt02	40548	11523	x	x
			FFPE_27912a	FFPE	Gt02	442396	331158	x	x
			27912b	Ultra II	Gt02	1145305	269073	x	x
			FFPE_27912b	FFPE	Gt02	4850675	808074	x	x

IV.C.1 hyRAD and ddRAD assembly

Combining the number of reads from the ddRAD, hyRAD and the concatenated FFPE-repaired and non-repaired individuals, 46 libraries with 93,686,915 reads were assembled by the mapping script (Section III.G.2.c.). The assembled data set consisted of 50,812,892 mapped reads, of which 5% were from the hyRAD individuals; and 365,451 SNPs were called by *freebayes*. After following the steps described in Section III.G.2.c., nine hyRAD libraries were excluded due to either more than 90% missingness or having a GC content outside the range of 40-47% (Fig. 9). Additionally, SNPs that were flagged as artefacts throughout the manual search in the Excel sheet, were removed. The final cleaned assembly contained 3,459 SNPs, consisting of five individuals from the 1960s, seven from the 1990s, four from the 2000s, and all recent Austrian and Slovakian individuals from the ddRAD assembly (Tab. 13). The subsequent filtering of the assembly's VCF file resulted in three different data sets with 606, 116, and 135 SNPs, respectively (Tab. 12).

Table 12: Number of SNPs retained for each hyRAD sub-assembly after read depth and SNP missing filtering.

Filtered hyRAD sub-assembly	Read depth (DP)	Allowed SNP missingness [%]	Nbr. of SNPs
1) hyRAD assembly sub-set	4	30	606
2) hyRAD assembly sub-set	4	25	116
3) hyRAD assembly sub-set	6	30	135

IV.C.2. Genetic diversity

The GC content of the hyRAD libraries varied more widely than that of the ddRAD individuals (39-55% vs. 41.86-42.34%, respectively) (Fig. 9). The oldest *L. apus* individuals from Marchegg (1961) had GC contents exceeding 47%; in contrast, 29127d from Marchegg Pvt (1992) had a GC content of less than 40%, while the FFPE-repaired and concatenated libraries of this individual showed GC contents of 41% and 40%, respectively. This consistency allowed 29127d to remain in the final hyRAD assemblies, whereas the other historic individuals were excluded. However, even without the outliers, the GC content of the remaining hyRAD libraries still showed considerable variation, ranging from 42% to 46%.

When comparing the FFPE-repaired, non-repaired and concatenated reads, the FFPE-repaired had the highest average GC content at 45%, while the non-repaired and concatenated libraries averaged 42% and 43%, respectively (Fig. 9).

Tab. 13: Summary of the ddRAD and hyRAD assemblies.

Note: Complete population names and individual ID prefixes are given in the figure text of Fig. 3.

Library preparation kits: FFPE: FFPE+G-tailing+Ultra II (Section III.C.1., Tab. 3), Ultra II: G-tailing+Ultra II (Section III.C.1., Tab. 3), FS: G-tailing+FS (Section III.C.1., Tab. 4), GA: Gansauge (Section III.C.2., Tab. 4), GAB: Gansauge with barcoded adapter (Section III:C:2., Tab. 4).

Note: Individual IDs starting with 'FFPE' correspond to extracts which were FFPE-repaired and starting with 'FP' correspond to concatenated reads from repaired and unrepaired extracts of the same individual.

** After removal of individuals with > 90% missingness and GC content < 40% or > 47% and excluding of non-ddRAD SNPs.*

Country	Population	Year	Individual ID	Library prep. kit	Library ID	Nbr. of mapped reads	Nbr. of SNPs post ddRAD loci filter	Nbr. of SNPs at DP 4 & 30% missing*	Nbr. of SNPs at DP 4 & 25% missing*	Nbr. of SNPs at DP 6 & 30% missing*
Austria	Marchegg	1961	FFPE_27707b	FFPE	Gt02	126	-	-	-	-
			27708	Ultra II	Gt02	76	-	-	-	-
	Marchegg Wgmld	1963	26712b	GA	GA1WDH	11114	1761	67	21	5
	Marchegg Pvt	1963	26718b	GA	GA1WDH	129643	3278	577	109	126
	Marchegg Pvt	1964	26719a	GA	Gt03	15390	2041	150	39	17
	Marchegg Wgmld	1965	26711b	GA	GA1WDH	18626	2137	157	32	22
	Parndorf	1966	27686b	GAB	GA1BWDH	11787	1741	83	16	10
	Marchegg Pvt	1990	27644a	GAB	GA1BWDH	1069	5	-	-	-
		1992	29127a	ddRAD (Section III.B.1.)	RAD02	6356417	3448	606	116	135
			29127b	ddRAD (Section III.B.1.)	RAD02	3388427	3442	606	116	135
			29127c	Ultra II	Gt02	146490	884	127	56	35
			FFPE_29127c	FFPE	Gt02	88752	401	64	41	28
			FP_29127c	concatenated	Gt02	235243	1105	170	56	48
			29127cc	Ultra II	Gt02	9611	89	-	-	-
			29127d	Ultra II	Gt02	143819	403	-	-	-
			FFPE_29127d	FFPE	Gt02	133838	391	48	31	15
			FP_29127d	concatenated	Gt02	277769	710	78	36	22
			29127e	Ultra II	Gt02	128984	596	-	-	-

			FFPE_29127e	FFPE	Gt02	2965	26	-	-	-
			FP_29127e	concatenated	Gt02	131943	615	102	35	32
			29127ee	GA	GA1WDH	52373	889	210	42	45
			29127f	FS	Gt03	5416	8	-	-	-
	Marchegg nW	1995	27677	GA	GA1WDH	18325	2090	170	42	31
	Marchegg TrSe	2000	27913a	ddRAD	RAD02	2034769	3417	599	113	133
			27913b	ddRAD	RAD02	1874399	3401	596	112	134
	Baumgarten	2002	27912a	Ultra II	Gt02	6618	55	-	-	-
			FFPE_27912a	FFPE	Gt02	82486	480	83	39	22
			FP_27912a	concatenated	Gt02	89110	514	87	42	25
			27912b	Ultra II	Gt02	125810	582	86	49	32
			FFPE_27912b	FFPE	Gt02	311665	808	107	51	34
			FP_27912b	concatenated	Gt02	437361	1177	120	77	50
	Austria	2013	30089	ddRAD (Section III.B.1.)	RAD15	2622813	3444	606	116	135
	Marchegg StW	2023	28383	ddRAD (Section III.B.1.)	RAD01	5446423	3431	605	116	134
	Marchegg Pvt	2023	28386	ddRAD (Section III.B.1.)	RAD01	3565294	3406	605	116	134
	Rabensburg	2023	28389b	ddRAD (Section III.B.1.)	RAD01	4801677	3442	605	116	135
			28389c	ddRAD (Section III.B.1.)	RAD01	2299736	3448	605	116	135
Slovakia	SVK west	2021	29700b	ddRAD (Section III.B.1.)	RAD11	398912	3320	585	115	134
			29700c	ddRAD (Section III.B.1.)	RAD11	725754	3429	606	116	135
			29700d	ddRAD (Section III.B.1.)	RAD12	526737	3416	604	116	134
	SVK east	2023	29701b	ddRAD (Section III.B.1.)	RAD12	3657355	3429	605	115	135
			29702c	ddRAD (Section III.B.1.)	RAD11	696969	3429	606	116	135
			29702f	ddRAD (Section III.B.1.)	RAD12	494035	3419	605	115	134
			29705d	ddRAD (Section III.B.1.)	RAD11	170124	3190	574	109	124
			29705e	ddRAD (Section III.B.1.)	RAD12	317811	3190	595	114	134

	SVK south	2023	29703	ddRAD (Section III.B.1.)	RAD11	182552	3351	593	109	122
			29704	ddRAD (Section III.B.1.)	RAD11	8636288	3426	605	115	135

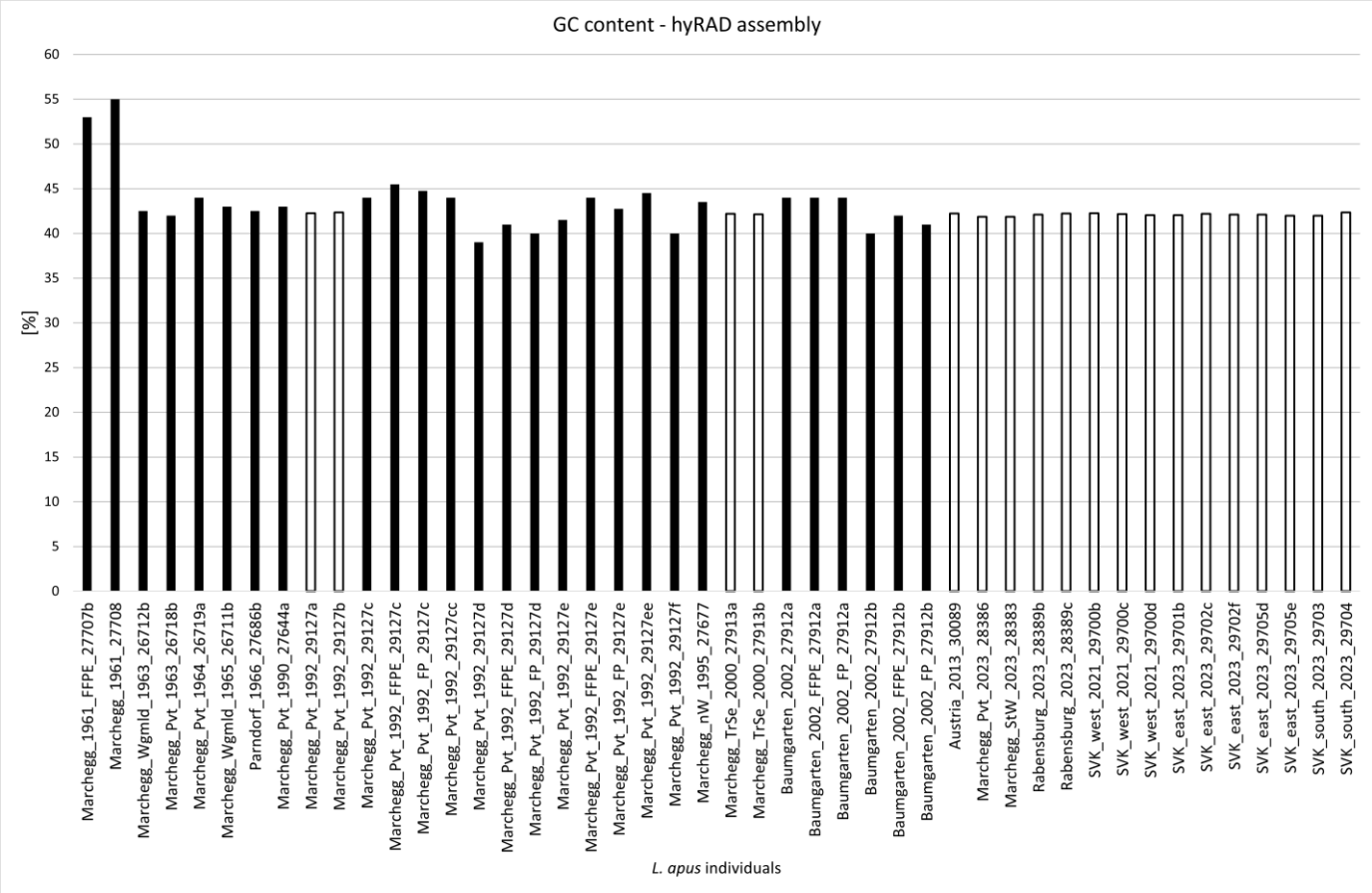


Figure 9: Bar plot showing the GC content of mapped reads of each *L. apus* individual from the hyRAD assembly.
White bars indicate the GC content of individuals originally from the ddRAD data.

The observed (H_O) values were calculated based on the SNPs of each hyRAD sub-assembly (Tab. 14).

The first sub-assembly (DP = 4, 30% missingness, 606 SNPs) showed H_O values ranging from 0 (Tab. 14, FFPE_ & FP_27912a) to 0.3718 (Tab. 14, 26711b). The highest H_O was observed in individuals from the 1960s, with an average H_O of 0.2778, particularly for 26711b (Marchegg Wgmld, 1965) and 26718b (Marchegg Pvt, 1963) (Tab. 14). High H_O values were overall visible for all individuals from the 1960s, with an average H_O of 0.2778 (Tab. 14). In contrast, individuals from the 1990s, 2002, and recent Austrian samples had lower H_O values (0.0575, 0.0463, and 0.0216, respectively). Slovakian individuals showed slightly higher H_O at 0.0532 (Tab. 14).

In the second hyRAD sub-assembly (DP = 4, 25% missingness, 116 SNPs), the average H_O for the 1960s individuals remained similar at 0.2636. The highest H_O was found in 26712b (Marchegg Wgmld, 1963) at 0.3333, while the lowest was 0.0 for FFPE-repaired 27912a (Baumgarten, 2002). Unlike the first assembly, H_O values increased from the 1990s to 2002 (0.0609 to 0.0774) (Tab. 14). Recent Austrian individuals had a higher H_O (0.04) compared to the previous sub-assembly, with Slovakian individuals showing an average H_O of 0.0571. Again, no obvious differences in H_O were noted between the FFPE-repaired, non-repaired, and concatenated sequences.

For the third hyRAD sub-assembly (DP = 6, 30% missingness, 135 SNPs), individual 26711b again had the highest H_O at 0.4091, while more individuals showed H_O values of 0.0 (e.g., FP_27912a, FFPE_29127d, 26712b) (Tab. 14). The H_O decreased similarly to the first sub-assembly: 1960s (H_O = 0.2525), 1990s (H_O = 0.0681), 2002 (H_O = 0.0585), and recent Austrian individuals (H_O = 0.018). The Slovakian individuals averaged H_O = 0.0467. Minimal differences were found between FFPE-repaired, non-repaired, and concatenated sequences, except for FFPE-repaired 29127d (H_O = 0.0) compared to its concatenated version (H_O = 0.05) (Tab. 14).

Table 14: Summary of SNP heterozygosity to represent genetic diversity of each *L. apus* individual from the different hyRAD sub-assemblies.

Note: Complete population names are given in the figure text of Fig. 3.

Note: Individuals <u>underlined</u> are originally from the ddRAD data.

Note: Individual IDs starting with 'FFPE' correspond to extracts which were FFPE-repaired and starting with 'FP' correspond to concatenated reads from repaired and unrepaired extracts of the same individual.

Country	Population	Year	Individual ID	Obs. Het (H_O) at DP 4 & 30% missing	Obs. Het (H_O) at DP 4 & 25% missing	Obs. Het (H_O) at DP 6 & 30% missing
Austria	Marchegg Wgmld	1963	26712b	0.2985	0.3333	0.0000
	Marchegg Pvt	1963	26718b	0.3183	0.2963	0.3360
	Marchegg Pvt	1964	26719a	0.1933	0.2821	0.1176
	Marchegg Wgmld	1965	26711b	0.3718	0.2813	0.4091
	Parndorf	1966	27686b	0.2073	0.1250	0.4000
	Marchegg Pvt	1992	29127a	0.0266	0.0348	0.0075
			29127b	0.0249	0.0435	0.0150
			29127c	0.0240	0.0364	0.0303
			FFPE_29127c	0.0317	0.0250	0.0370
			FP_29127c	0.0357	0.0290	0.0217
			FFPE_29127d	0.0217	0.0333	0.0000
			FP_29127d	0.0267	0.0571	0.0500
			FP_29127e	0.0396	0.0882	0.1290
			29127ee	0.0478	0.0238	0.0682
	Marchegg nW	1995	27677	0.2959	0.2381	0.3226
	Marchegg TrSe	2000	27913a	0.0202	0.0625	0.0305
			27913b	0.0270	0.0450	0.0152
	Baumgarten	2002	FFPE_27912a	0.0000	0.0000	0.0000
			FP_27912a	0.0000	0.0000	0.0000
			27912b	0.0595	0.0833	0.0667
			FFPE_27912b	0.1154	0.1800	0.1515
			FP_27912b	0.1018	0.1711	0.1458
	Austria	2013	30089	0.0299	0.0435	0.0150
	Marchegg StW	2023	28383	0.0133	0.0435	0.0150
	Marchegg Pvt	2023	28386	0.0166	0.0348	0.0150
	Rabensburg	2023	28389b	0.0233	0.0435	0.0226
			28389c	0.0250	0.0348	0.0226
Slovakia	SVK west	2021	29700b	0.0413	0.0439	0.0303
			29700c	0.0316	0.0522	0.0226
			29700d	0.2333	0.2000	0.2045
	SVK east	2023	29701b	0.0250	0.0351	0.0226
			29702c	0.0299	0.0435	0.0150
			29702f	0.0366	0.0526	0.0530
			29705d	0.0385	0.0278	0.0488
			29705e	0.0304	0.0354	0.0227
	SVK south	2023	29703	0.0356	0.0370	0.0248
			29704	0.0300	0.0435	0.0226

IV.C.3. Population structure

In the first sub-assembly (DP = 4, 30% missingness, 606 SNPs), eastern and southern Slovakian individuals were clearly separated from the western and Austrian groups, with the exception

of 29701b, which clustered closely with the historic individual 26718b (Marchegg Pvt, 1963) (Fig. 10b). This overall clustering pattern was consistent with the ddRAD assembly PCA, where west Slovakian individuals grouped near recent Austrian samples (Fig. 6). Once the eastern and southern Slovakians were removed, 26718b appeared furthest from the remaining individuals (Fig. 10c, e). The two Marchegg TrSe individuals (2000) and 29127a/29127b (Marchegg Pvt, 1992) were separated from other Austrian individuals, but the majority of the Marchegg Pvt individuals (1992) clustered with those from the 1960s, including 27686b from Parndorf (Fig. 10b, c). Individuals 26719a (Marchegg Pvt, 1964) and 27677 (Marchegg nW, 1995) appeared more distinct from the central cluster.

The general pattern of eastern and southern Slovakian individuals separating from the western and Austrian groups was again apparent in the second hyRAD sub-assembly (DP = 4, 25% missingness, 116 SNPs) (Fig. 11). A notable exception was 27913a (Marchegg TrSe, 2000), which was as distant from the Austrian cluster as the eastern and southern Slovakian individuals (Fig. 11b, c), while the second individual from this population was positioned closer to the centre (Fig. 11c). The previously observed separation of 29127a/29127b (Marchegg Pvt, 1992) from the other Austrians, as seen in the first sub-assembly and the ddRAD PCA, shifted closer to the Austrian cluster when eastern/southern Slovakians were included (Fig. 11b, d), but reappeared once they were excluded. Interestingly, 26718b was no longer furthest from the group, but within the Austrian cluster (Fig. 11). In the fourth PCA, the historic individuals spread more widely than in previous sub-assemblies, with Marchegg Pvt individuals from 1992 mixing with early 2000s and 28389c from Rabensburg (2023) (Fig. 11e).

The PCA plots of the third sub-assembly (DP = 6, 30% missingness, 135 SNPs) supported the observations of the first sub-assembly PCAs (Fig. 12). The ddRAD individuals were clustered around the hyRAD individuals, and no obvious outliers were detected. However, 29127a/b (Marchegg Pvt, 1992) clustered closely with west Slovakian individual 29700b when the eastern and southern Slovakian individuals were excluded (Fig. 12c, e). 26718b remained aligned with the Austrian individuals, similar to the second hyRAD sub-assembly.

Regarding the FFPE-repaired, non-repaired, and concatenated samples (e.g., 27912b from Baumgarten, 2002), they clustered closely together with no noticeable outliers (Fig. 10, 11, 12).

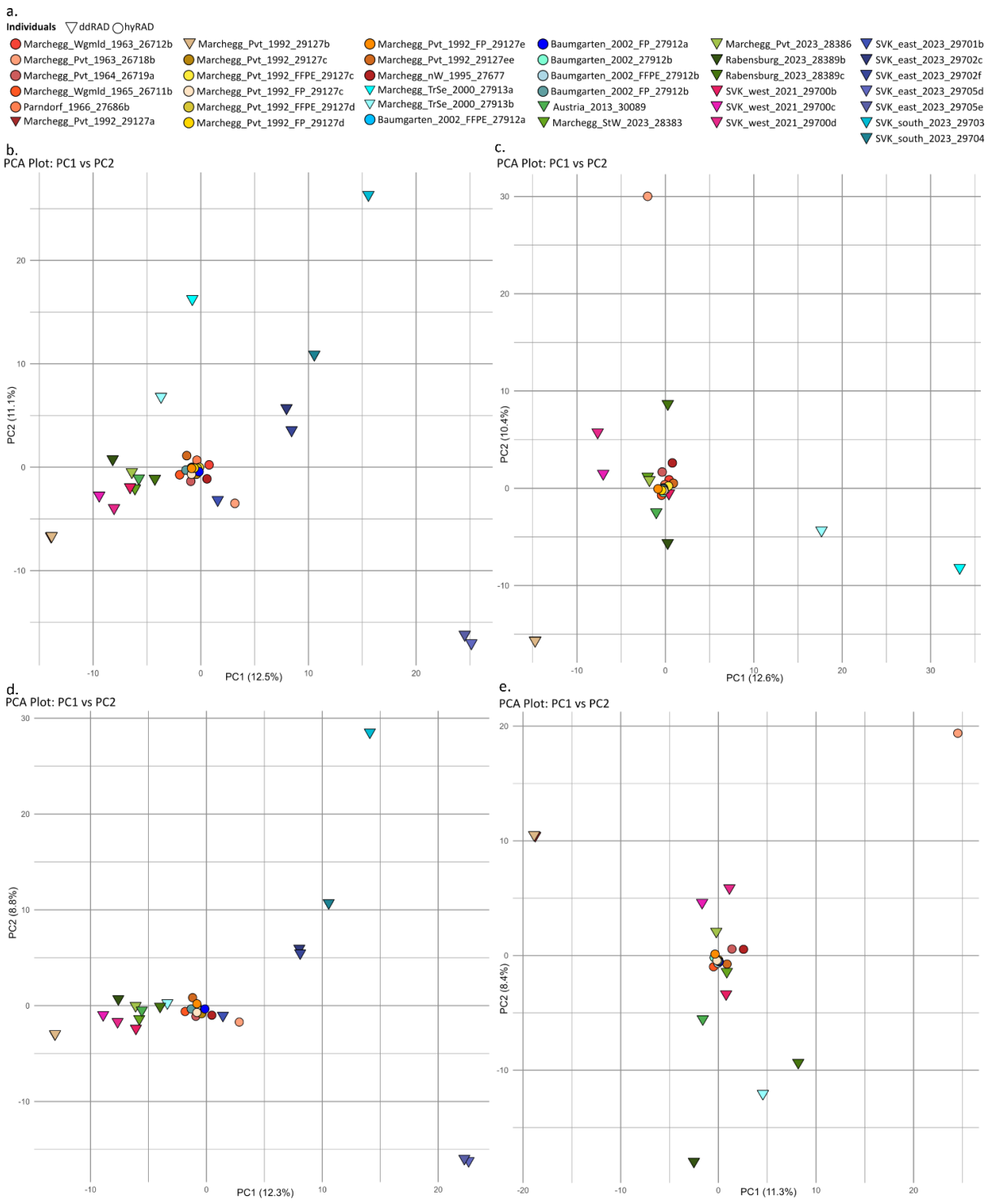


Figure 10: Analysis of principal component (PC) plots for the 1) hyRAD sub-assembly with DP = 4, 30% missingness, and 606 SNPs. a. Legend of ddRAD (triangle) and hyRAD (circle) individuals in PCA plots (b-e). **b.** PCA plot with all individuals of sub-assembly. **c.** PCA plot excluding eastern and southern Slovakian individuals. **d.** PCA plot excluding historic individuals with > 85% missingness. **e.** PCA plot excluding eastern/southern Slovakian individuals, and historic individuals with > 85% missingness. See Fig. S4 for scree plots.

Figure 11: Analysis of principle component (PC) plots for the 2) hyRAD sub-assembly with DP = 4, 25% missingness, and 116 SNPs. a. Legend of ddRAD (triangle) and hyRAD (circle) individuals in PCA plots (b-e). **b.** PCA plot with all individuals of sub-assembly. **c.** PCA plot excluding eastern and southern Slovakian individuals. **d.** PCA plot excluding historic individuals with > 85% missingness. **e.** PCA plot excluding eastern/southern Slovakian individuals, and historic individuals with > 85% missingness. See Fig. S4 for scree plots.

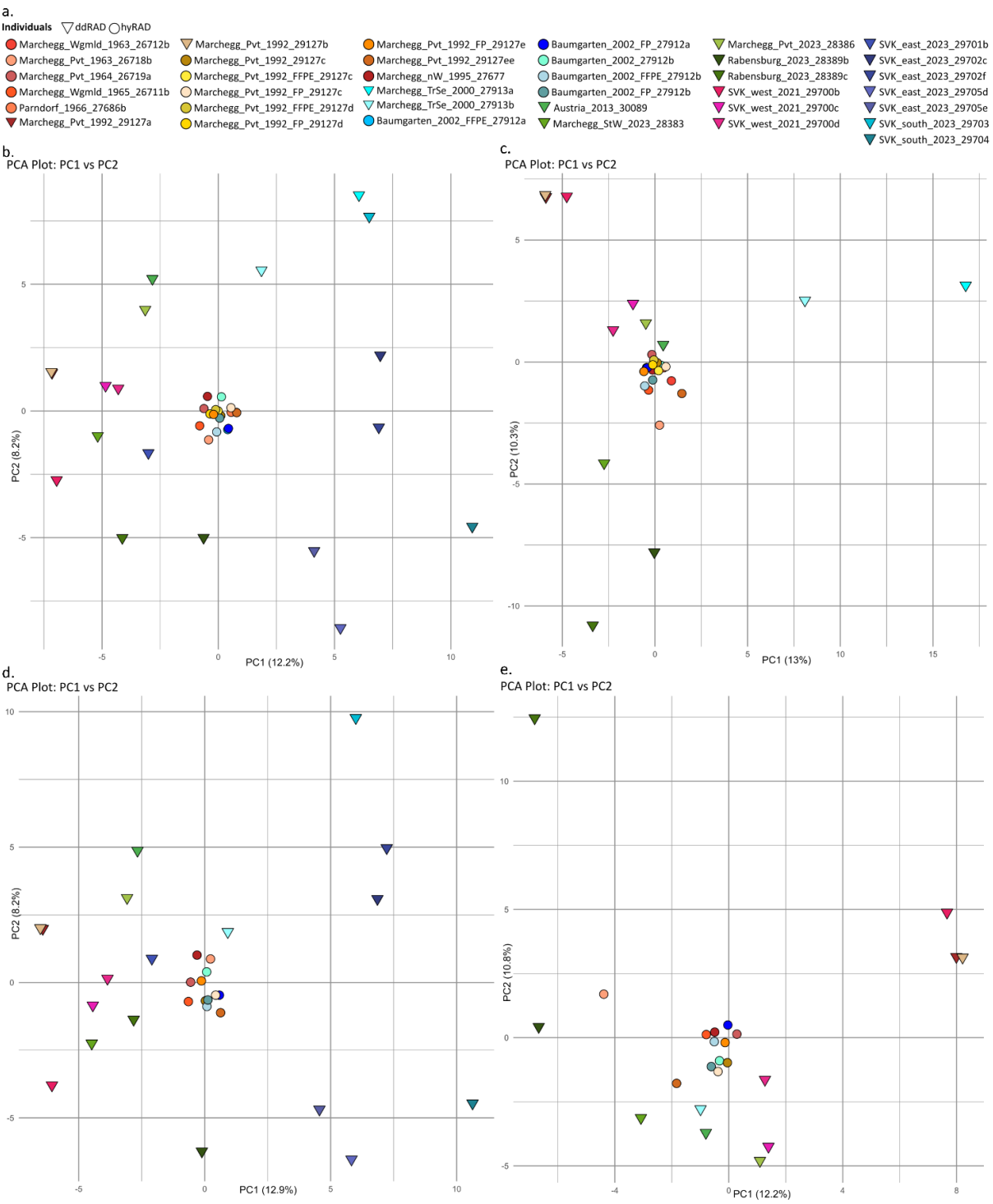


Figure _: Analysis of principle component (PC) plots for the 3) hyRAD sub-assembly with DP = 6, 30% missingness, and 135 SNPs. a. Legend of ddRAD (triangle) and hyRAD (circle) individuals in PCA plots (b-e). **b.** PCA plot with all individuals of sub-assembly. **c.** PCA plot excluding eastern and southern Slovakian individuals. **d.** PCA plot excluding historic individuals with > 85% missingness. **e.** PCA plot excluding eastern/southern Slovakian individuals, and historic individuals with > 85% missingness. See Fig. S4 for scree plots.

V. DISCUSSION

V.A. Comparison of ssDNA library preparation methods

In this Thesis, the comparison of two ssDNA library preparation protocols highlighted practical differences. The *G-tailing* method (Tin et al., 2014) produced on average more quality reads than the *Gansauge* method (Gansauge and Meyer, 2013; Gansauge et al., 2020). This was unexpected as the concentration of the historic input DNA was generally low and therefore more suitable for *Gansauge* than for the *G-tailing* protocol (Gansauge and Meyer, 2013; Tin et al., 2014). However, the *Gansauge* approach is more complex in laboratory handling, particularly the adapter ligation and cleanup steps, which could have led to errors and DNA loss (Kapp et al., 2021). This may have contributed to the lower sequencing success in this study. Additionally, Gansauge et al. (2020) noted that library quality decreases with input DNA fragments > 250 bp, which might have been the case for some historic *L. apus* individuals. While DNA fragment lengths were assessed for some individuals using the 4200 TapeStation System, with an average length of 200 bp observed, further fragment length assessments for the remaining historic individuals are needed for a more thorough assessment.

The results for the FFPE-repaired DNA extracts were mixed. Although the repaired libraries initially produced on average more raw reads than the non-repaired ones, after quality filtering fewer reads on average were retained. These results are in contrast to a comparative approach to generate COI mini-barcodes from FFPE-repaired and non-repaired historic isopods (Barta, 2023). There, the success rate of amplifying the COI region was higher in FFPE individuals, likely due to the repair of DNA single-strand breaks and crosslinks. This suggests that herein the FFPE repair process worked less effectively than anticipated.

Given the results, the *G-tailing* protocol was more effective as it provided more quality reads on average and a better laboratory handling. Its ability to handle a broader range of input DNA concentrations (Tin et al., 2014), and more modest protocol, especially timewise, made it more practical for processing the larger number of historic DNA extracts involved in this study. The high number of cycles for the index PCR (25 cycles, Tab. 3, Step 5) did not greatly reduce the quantity or quality of reads, which contrasts with the *Gansauge* workflow's lower output despite the fewer PCR cycles (20 cycles, Tab. 4, Step 11).

As an alternative to both ssDNA library preparation methods, Kapp et al. (2021) present the *Santa Cruz Reaction* (*SCR*). Like the *G-tailing*, *SCR* is more time-efficient while providing sufficient results for degraded samples at a low DNA input, comparable to *Gansauge* (Kapp et al., 2021). This could potentially be a valuable solution for future work with historical or degraded material, combining the time efficiency of *G-tailing* with the accuracy of ssDNA recovery in the *Gansauge* protocol.

V.B. *Lepidurus apus* genome assembly

Four *Lepidurus apus* genomes were assembled for this study, and the most optimal assembly (Lapus_Atr) was selected as reference for the ddRAD and hyRAD assemblies. The genome, successfully assembled using Illumina short reads from a single individual collected in Rabensburg (Austria) 2023, was chosen due to its favourable genome statistics. This approach was similarly applied in the assemblies of *L. apus*, *L. couesii* (Luchetti et al., 2021), *L. arcticus* and *L. lubbocki* (Savojardo et al., 2019). In contrast, the assembly of *L. packardi* was done using a combination of PacBio long reads and Illumina short reads (Blair et al., 2022a). Despite relying solely on short-read sequencing, the selected assembly for this study resulted in a *L. apus* genome of 84.3 Mb (16,810 contigs) which is consistent with the known tendency for Notostraca species to have smaller genomes compared to other Crustaceans (Jeffery, 2015; Savojardo et al., 2019; Luchetti et al., 2021; Blair et al., 2022a).

The Lapus_L and Lapus_AL assemblies had more reads and higher coverage but produced longer genomes than Lapus_Atr, despite smaller estimated genome sizes. These results are not consistent with the genome sizes known for other *Lepidurus* species such as *L. couesii* (74.1 Mb) (Luchetti et al., 2021), *L. arcticus* (73.2 Mb) (Savojardo et al., 2019), or *L. lubbocki* (90.3 Mb) (Savojardo et al., 2019). Both assemblies also had more contigs and lower N50 values than Lapus_Atr, suggesting potential overassembly or presence of unnecessary sequences that should have been removed prior to or after the assembly.

While the Lapus_A assembly had a similar N50 value, its higher contig count made Lapus_Atr the more qualified choice.

Luchetti et al. (2021) assembled an even smaller *L. apus* genome, with a total length of 78 Mb and 4,988 contigs, achieving a 98% *BUSCO* completeness. The Lapus_Atr genome of this study is slightly larger than that reported by Luchetti et al. (2021) for *L. apus*, but the estimated genome size (83.3 Mb) is closer, and the *BUSCO* completeness of 97.3% suggests that this assembly is reliable as a reference genome. Larger genomes have been described for *L. lubbocki* (90.3 Mb) (Savojardo et al., 2019) and *L. packardi* (108.6 Mb) (Blair et al., 2022a), showing a broad range of genome sizes within the *Lepidurus* genus.

The GC content of the *L. apus* genome presented here (42.06%) is within the range observed for the other *Lepidurus* genomes, such as *L. apus* (41.5%), *L. couesii* (41.7%) (Luchetti et al., 2021), and *L. packardi* (40.9%) (Blair et al., 2022a). This similarity extends to two other Notostracans, *Triops longicaudatus* with 40.2% and *T. cancriformis* with 41.3% (Luchetti et al., 2021). This GC content appears to be similar across large Branchiopoda, as observed in the Anostracans *Branchinecta lynchi* (41.4%) (Blair et al., 2022c) and *B. lindahli* (41.9%) (Blair et al., 2022b).

Lepidurus apus shows low heterozygosity (0.05%), similar to *L. arcticus* (0.03%) (*Savojardo et al., 2019*), likely due to reproductive strategies like parthenogenesis or hermaphroditic self-fertilising, which limit genetic diversity (*Longhurst, 1955a, 1955b; Scanabissi and Mondini, 2002; Savojardo et al., 2019; Blair et al., 2022a*). In contrast, *L. packardi* has a higher heterozygosity of 0.79%, and while the alternative reproductive modes of this species' populations are believed to be fully bisexual, fully hermaphroditic, or mixed, they remain largely unknown (Blair et al., 2022a). Other large Branchiopoda, such as *Branchinecta lynchi* (1.98%) and *B. lindahli* (2.85%) (*Blair et al., 2022c, 2022b*), display even higher levels. *Eulimnadia texana*, a hermaphroditic clam shrimp, concentrates 64% of its total heterozygosity in just 5.6 Mb of its 120 Mb genome (*Baldwin-Brown et al., 2018*). Overall, *L. apus* stands out as having particularly low genomic heterozygosity compared to other large Branchiopoda species.

V.C. RAD data

V.C.1. Challenges in artefactual SNP detection

In both the ddRAD and hyRAD assemblies, detecting artefactual SNPs in the historic libraries posed a great challenge. Initial SNP calling by *ipyrad* for the ddRAD assembly resulted in 5,163 SNPs, 289 of which were later identified as artefacts, appearing exclusively in single historic individuals. This identification became clearer toward the end of the study, following a detailed analysis of the hyRAD assemblies.

In the hyRAD data, the artefacts were detected by comparing SNPs between FFPE-repaired and non-repaired library replicates from the same individual, revealing SNPs present in one replicate but not the other or any other library in the assembly. These false SNPs had sufficient read support and would not have been identified without replicates, emphasising the importance of using replicates for effective artefact detection in historic hyRAD libraries.

Artefactual SNPs observed in both ddRAD and hyRAD assemblies were consistent with findings in previous studies, where more artefacts were detected in historic libraries due to historic DNA (Suchan et al., 2016; Linck et al., 2017; Crates et al., 2019; Lavretsky et al., 2019; Takashi et al., 2019; Gauthier et al., 2020; Masharing et al., 2023). introduced errors are commonly attributed to DNA fragmentation, chemically induced modifications, and contamination in older samples.

V.C.1.a. Base modifications in historic DNA

Typical base modifications in degraded DNA, such as C-to-T transitions, occur post-mortem and are often reinforced by chemical treatments like formalin fixation (Do and Dobrovic, 2014; Suchan et al., 2016). Although several formalin-fixed hyRAD libraries in this study (e.g., 26718b, 26719a, 27686b; SI CM Tab.) showed no notably low GC content, the oldest, formalin-fixed libraries, FFPE_27707b and 27708 from Marchegg (1961), exhibited higher-than-average GC content. Both were prepared using the *G-tailing* protocol, and despite thorough quality checks of the already small number of suitable reads, the trimming may have been insufficient. These libraries were included in the hyRAD assembly mainly for their historical value. Nonetheless, due to their GC content and high missingness, both libraries were excluded in the hyRAD sub-assemblies. Interestingly, library 29127d (Marchegg Pvt, 1992) had a GC

content of 39%, while the FFPE-repaired version had 41%, possibly indicating successful cytosine deamination repair. In general, FFPE-repaired libraries showed higher GC content than unrepaired ones, suggesting that the repair enzymes may have corrected naturally occurring 'T' bases rather than those from deamination. However, this would need further investigation at the nucleotide level.

Overall, the hyRAD libraries showed greater GC content variation than the ddRAD libraries, as seen in other hyRAD studies (Crates et al., 2019; Gauthier et al., 2020). This variation may result from the notably fewer loci captured in the hyRAD libraries, making the GC content more susceptible to statistical deviations depending on which loci are included by chance.

To address base modifications, post-mortem DNA correction tools like *mapDamage2.0* (Jónsson et al., 2013) are available. In this study, strict filtering was deemed adequate to minimise such impacts (Suchan et al., 2016; Linck et al., 2017; Crates et al., 2019; Gauthier et al., 2020).

V.C.1.b. Impact of PCR amplifications on SNP accuracy

Errors and amplification biases from PCR are common sources of artefacts in both ddRAD and hyRAD data, particularly when amplifying degraded DNA (Linck et al., 2017; Lavretsky et al., 2019; Takashi et al., 2019; Masharing et al., 2023). The historic ddRAD library (library ID: RAD02) and the hyRAD libraries were amplified using 15 cycles. While increasing PCR cycles could have resulted in more reads, it likely would not reduce artefactual SNPs and may worsen the issue (Takahashi et al., 2019). To prevent the influence of error in the ddRAD data, PCR duplicates were removed prior to the assembly (Linck at al., 2017). However, removing duplicates further limits SNPs for analysis in degraded samples, which is why in this study, PCR duplicates were kept in hyRAD data to maintain sufficient coverage.

V.C.1.c. Thresholds in SNP calling

Setting appropriate SNP calling thresholds could have helped reduce artefacts in historic ddRAD data (e.g., setting a minimum of 30% for observed alleles, as used in the hyRAD assembly mapping) (Lavretsky et al., 2019). However, the degraded state of historic DNA often leads to low coverage and higher missing data, making SNP detection challenging even when using more stringent filters (Lavretsky et al., 2019; Takashi et al., 2019).

V.C.1.d. Limitations of read depth in historic SNP detection

Insufficient read depth, particularly in historic libraries, can lead to introducing artefactual SNPs into ddRAD and hyRAD assemblies. Masharing et al. (2023) emphasised the importance of adequate read depth for reliable SNP calling. A read depth of 6 in the ddRAD assembly appeared sufficient for recent libraries but may not have been for historic ones. Some libraries showed higher read counts for incorrect alleles than correct ones (sometimes exceeding them by a factor of two), suggesting that increased coverage might not have prevented artefacts. In hyRAD assemblies, similar challenges were observed. Crates et al. (2019) detected more artefactual ('private') SNPs in historic populations, noting that the poorer quality of historic DNA likely introduced errors. Even after increasing the minimum DP to 15, artefactual SNPs remained. Strict filtering remains necessary but challenging in degraded DNA studies (Linck et al., 2017; Gauthier et al., 2020).

V.C.2. SNP comparison in ddRAD and hyRAD assemblies

V.C.2.a. Comparison of uncorrected and corrected ddRAD assemblies

Comparing PCA outcomes from uncorrected (5,163 SNPs) and corrected (4,874 SNPs) ddRAD assemblies showed a similar genetic pattern, with a major division between Austrian/west Slovakian and east/south Slovakian groups (Fig. 8). However, the uncorrected PCA presented a more fragmented structure, particularly concerning the clustering of the historic individuals. Recalculating the PCA with corrected ddRAD assembly data revealed closer clustering of the historic with recent individuals, reducing apparent genetic distance. This suggests that removed SNPs had artificially increased the genetic distances, misrepresenting the true population structure.

Subsequent discussions on population structure are solemnly based on the PCAs of the corrected ddRAD assembly, excluding results which were based on the uncorrected ddRAD assembly, namely the Neighbour-net network, STRUCTURE analysis, and fineRADstructure. With regard to the observed heterozygosity, only the values of the recent individuals are discussed.

V.C.2.b. Comparison of hyRAD sub-assemblies

The clustering of historic individuals in the hyRAD sub-assemblies showed low variance across all PCAs, likely due to missing data and low coverage. Despite limitations, the same broad clustering pattern emerged in all three sub-assemblies, matching ddRAD PCA's division between Austrian/west Slovakian and east/south Slovakian groups. The consistency suggests that, at a broad level, the sub-assemblies captured the key population structure regardless of the filtering specifics.

The second sub-assembly appears to be the least reliable due to its lower SNP count and broader spread of historic individuals, which may have compromised resolution. Both the first and third sub-assemblies provided more robust results. The first retained more SNPs, while the third, with fewer SNPs, but better read depth, might offer a more accurate representation of genetic relationships.

Further population structure analyses are needed to confirm which sub-assembly best represents the genetic relationships. Therefore, as the decision remains inconclusive and the historic hyRAD data is affected by low coverage and high missingness, the subsequent discussion on temporal patterns of Austrian populations must be approached with the utmost care.

V.C.3. Genetic diversity of recent *L. apus* individuals from Austria and Slovakia

The observed heterozygosity (H_O) in recent *Lepidurus apus* individuals was low overall, with Slovakian individuals showing slightly higher values than Austrian ones. This pattern likely reflects the reproductive strategy of *L. apus*, where hermaphroditism might have contributed to reduced genetic diversity through self-fertilisation (Longhurst, 1955a, 1955b; Scanabissi and Mondini, 2002; Savojardo et al., 2019; Blair et al., 2022), even in the resting eggs, reducing their function to maintain long-term genetic diversity within populations and buffer against genetic variation (Evans and Dennehy, 2005; Mantovani et al., 2008; Schwentner and Richter, 2015). While short-distance dispersal mechanisms, such as wind and birds, may support genetic connectivity within the Austrian populations, these populations likely remain isolated on a broader geographic scale, particularly from eastern and southern Slovakian populations. Similar patterns were observed in other large Branchiopoda, such as the hermaphroditic

Triops cancriformis in Austria (Mantovani et al., 2008), or *Branchinecta sandiegonensis* in California (David et al., 1997) and *Leptestheria dahalacensis* in Europe, where low heterozygosity results from habitat fragmentation and limited passive dispersal (Cesari et al., 2007).

V.C.4. Population structure

V.C.4.a. Spatial population structure: *L. apus* within Austria and west Slovakia

The west Slovakian populations showed low genetic differentiation from the Austrian *L. apus*, particularly the Marchegg individuals, likely due to passive dispersal mechanisms and their geographical proximity. The Morava, separating eastern Austria from western Slovakia, may act as a dispersal route, with floodings and birds serving as natural vectors for resting egg dispersals (Fig. 8) (Bilton et al., 2001; De Meester et al., 2002; Cesari et al., 2007; Vanschoenwinkel et al., 2011). Wind may also contribute to maintaining close genetic relationships across the short distance between Austria and west Slovakia (Rodríguez-Flores et al., 2017). Similar mechanisms were observed in *Branchipus schaefferi*, which exhibited low genetic divergence across Poland, despite a wide geographical distribution of populations. This was attributed to passive distance egg dispersal and the absence of interfering geographic barriers (Mioduchowska et al., 2018).

The limited genetic diversity within the Austria/west Slovakia cluster may also be influenced by hermaphroditism through self-fertilisation (Longhurst, 1955a, 1955b; Scanabissi and Mondini, 2002; Savojardo et al., 2019; Blair et al., 2022). This has also been observed in other Notostraca species, such as *T. cancriformis*, where populations exhibiting hermaphroditism show lower genetic variability (Mantovani et al., 2004, 2008; Velonà et al., 2009).

The Rabensburg (2023) individuals form a genetically distinct subset within the Austrian/west Slovakian cluster. Their geographical separation from Marchegg (43 km) and proximity to the north Austrian Thaya River might have contributed to local genetic isolation (Fig. 8). With resting egg banks helping to preserve a distinct gene pool, this distance likely explains their more distant genetic relationship to the Marchegg and west Slovakian individuals (Evans and Dennehy, 2005; Cesari et al., 2007; Schwentner and Richter, 2015; Rodríguez-Flores et al., 2017; Sainz-Escudero et al., 2023).

V.C.4.b. Spatial population structure: *L. apus* between Austria/west Slovakia and east/south Slovakia

The genetic separation between Austrian/west Slovakian and east/south Slovakian *L. apus* populations suggests that the geographical distance is too great for effective long-distance passive dispersal by birds or wind, compounded by the lack of connected river systems which could aid as routes (Fig. 8). The Little Carpathians, a low mountain range in western Slovakia, might further act as a barrier to gene flow between eastern and southern Slovakia, reinforcing genetic separation. Comparable patterns of isolation have been observed in *Triops cancriformis,* where the Austrian Marchegg populations are genetically distant from the other European populations. This isolation may be due to both geographic distance and unique environmental conditions in Marchegg that could limit the establishment of individuals from other regions (Mantovani et al., 2008; Velonà et al., 2009). Geographical barriers like the Alps similarly reduce gene flow between Central European and Italian populations of Spinicaudata *Leptestheria dahalacensis* (Cesari et al., 2007), or mountain basins within the Iberian Peninsula that function as barriers in the Anostracans *Branchinecta ferox* (Rodríguez-Flores et al., 2017) and *Branchinectella media* (Sainz-Escudero et al., 2023).

The east and south Slovakian populations of *L. apus,* though forming a roughly coherent cluster, are apparent from each other, likely due to their considerable geographic distance (242 km). Given this distance, an even greater genetic difference might have been expected, as seen in *Branchinectella media,* where populations approximately 300 km apart, form distinct genetic clusters (Sainz-Escudero et al., 2023). Unlike *B. media*, which is further on separated by mountain terrain, the eastern and southern *L. apus* populations are not divided but flanked to the north by the Slovak Ore Mountains, potentially facilitating passive long-distance dispersal from south-west to east and influencing the genetic relationship of these geographically distant populations.

V.C.4.c. Temporal population structure: *L. apus* in Austria over the past 60 years

The comparison between historic (1960s, 1990s, and 2000s) and recent Austrian *L. apus* individuals suggests an initial genetic distinctiveness in older individuals that gradually transitioned into a more uniform genetic structure over time.

The unique cluster of the 1960s Marchegg individuals suggests some genetic differentiation, while the 1990s Marchegg individuals clustered closer to the recent ones, possibly indicating a gradual decrease in genetic differentiation over time. Between the 1990s Marchegg individuals and recent ones, reduced genetic distinctiveness suggests a gradual loss of unique genetic traits and increased genetic uniformity over the last 20 years. This uniformity potentially aligns with the clustering of the 2013 Austrian individual with recent individuals, indicating overall temporal genetic stability. Similar subtle temporal shifts with genetic stability were observed in the New Guinea kingfisher (*Syma torotoro*), which showed genetic continuity across samples spanning up to 140 years (Linck et al., 2017). The regent honeyeater *Anthochaera phrygia*, a nomadic Australian songbird, also displayed genetic stability despite habitat fragmentation, with both historic and recent populations forming a unified genetic group due to maintained gene flow over distance (Crates et al., 2019).

The noticeable clustering of the *L. apus* individuals from Baumgarten (2002), located 7 km from Marchegg, may be attributed to local genetic isolation over time (Zofkova, 2006; Gauthier et al., 2020; Lukić et al., 2019). Such patterns have been observed in *Erebia embla*, a butterfly species in which habitat fragmentation led to increased isolation and reduced genetic diversity over time (Gauthier et al., 2020). Additionally, resting egg banks may have preserved a distinct gene pool by limiting gene flow from external migration (De Meester et al., 2002; Evans and Dennehy, 2005; Cesari et al., 2007; Schwentner and Richter, 2015; Rodríguez-Flores et al., 2017; Sainz-Escudero et al., 2023).

In *L. apus*, short-distance dispersal by water, wind and birds may have maintained genetic connectivity across neighbouring ponds, helping stabilise genetic diversity in the Austrian Marchegg populations.

VI. LITERATURE

Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. (n.d.). http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Baldwin-Brown, J. G., Weeks, S. C., & Long, A. D. (2017). A New Standard for Crustacean Genomes: The Highly Contiguous, Annotated Genome Assembly of the Clam Shrimp *Eulimnadia texana* Reveals HOX Gene Order and Identifies the Sex Chromosome. *Genome Biology and Evolution*, *10*(1), 143–156. https://doi.org/10.1093/gbe/evx280

Barta, A.-C. (2023). DNA barcoding of terrestrial isopods in Austria integrating museum specimens [Master's Thesis]. University of Vienna, Vienna, Austria.

Bilton, D. T., Freeland, J. R., & Okamura, B. (2001). Dispersal in aquatic invertebrates. *Annual Revision Ecological Systems*, *32*, 159–181.

Blair, S. R. K., Hull, J., Escalona, M., Finger, A., Joslin, S. E. K., Sahasrabudhe, R., Marimuthu, M. P. A., Nguyen, O., Chumchim, N., Morris, E. R., Velazquez, S., & Schreier, A. (2022a). The reference genome of the Vernal Pool Tadpole Shrimp, *Lepidurus packardi*. *Journal of Heredity*, *113*(6), 706–711. https://doi.org/10.1093/jhered/esac051

Blair, S. R. K., Schreier, A., Escalona, M., Finger, A. J., Joslin, S. E. K., Sahasrabudhe, R., Marimuthu, M. P. A., Nguyen, O., Chumchim, N., Morris, E. R., Mangelson, H., & Hull, J. (2022b). A chromosome-level reference genome for the Versatile Fairy Shrimp, *Branchinecta lindahli*. *Journal of Heredity*, *114*(1), 74–80. https://doi.org/10.1093/jhered/esac057

Blair, S. R. K., Schreier, A., Escalona, M., Finger, A. J., Joslin, S. E. K., Sahasrabudhe, R., Marimuthu, M. P. A., Nguyen, O., Chumchim, N., Morris, E. R., Mangelson, H., & Hull, J. (2022c). A draft reference genome of the Vernal Pool Fairy Shrimp, Branchinecta lynchi. *Journal of Heredity*, *114*(1), 81–87. https://doi.org/10.1093/jhered/esac056

Brendonck, L. (1996). Diapause, quiescence, hatching requirements: what we can learn from large freshwater branchiopods (Crustacea: Branchiopoda: Anostraca, Notostraca, Conchostraca). *Hydrobiologia*, *320*, 85–97.

Brendonck, L., & De Meester, L. (2003). Egg banks in freshwater zooplankton: evolutionary and ecological archives in the sediment. *Hydrobiologia*, *491*(1–3), 65–84. https://doi.org/10.1023/a:1024454905119

Burrell, A. S., Disotell, T. R., & Bergey, C. M. (2014). The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution*, *79*, 35–44. https://doi.org/10.1016/j.jhevol.2014.10.015

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*(1). https://doi.org/10.1186/1471-2105-10-421

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and Genotyping LOCI de novo from Short-Read Sequences. *G3 Genes Genomes Genetics*, *1*(3), 171–182. https://doi.org/10.1534/g3.111.000240

Cesari, M., Luchetti, A., Scanabissi, F., & Mantovani, B. (2007). Genetic variability in European *Leptestheria dahalacensis* (Rüppel, 1837) (Crustacea, Branchiopoda, Spinicaudata). *Hydrobiologia*, *586*(1), 249–260. https://doi.org/10.1007/s10750-007-0645-2

Challis, R., Richards, E., Rajan, J., Cochrane, G., & Blaxter, M. (2020). BlobToolKit – Interactive Quality Assessment of Genome Assemblies. *G3 Genes Genomes Genetics*, *10*(4), 1361–1374. https://doi.org/10.1534/g3.119.400908

Crates, R., Olah, G., Adamski, M., Aitken, N., Banks, S., Ingwersen, D., Ranjard, L., Rayner, L., Stojanovic, D., Suchan, T., Von Takach Dukai, B., & Heinsohn, R. (2019). Genomic impact of severe population decline in a nomadic songbird. *PLoS ONE*, *14*(10), e0223953. https://doi.org/10.1371/journal.pone.0223953

Damgaard, J., & Olesen, J. (1998). Distribution, phenology and status for the larger branchiopoda (Crustacea: Anostraca, Notostraca, Spinicaudata and Laevicaudata) in Denmark. *Hydrobiologia*, *377*(1/3), 9–13. https://doi.org/10.1023/a:1003256906580

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2). https://doi.org/10.1093/gigascience/giab008

De Meester, L., Gómez, A., Okamura, B., & Schwenk, K. (2002). The monopolization hypothesis and the dispersal-gene flow paradox in aquatic organisms. *Acta Oecologica: International Journal of Ecology*, *23*, 121–135.

Derkarabetian, S., Benavides, L. R., & Giribet, G. (2019). Sequence capture phylogenomics of historical ethanol-preserved museum specimens: Unlocking the rest of the vault. *Molecular Ecology Resources*, *19*(6), 1531–1544. https://doi.org/10.1111/1755-0998.13072

Do, H., & Dobrovic, A. (2014). Sequence Artifacts in DNA from Formalin-Fixed Tissues: Causes and Strategies for Minimization. *Clinical Chemistry*, *61*(1), 64–71. https://doi.org/10.1373/clinchem.2014.223040

Eaton, D. a. R., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, *36*(8), 2592–2594. https://doi.org/10.1093/bioinformatics/btz966

Eder, E., & Hödl, W. (2002). Large freshwater branchiopods in Austria: diversity, threats and conservational status. *Escobar-Briones & Alvarez, Eds., Modern Approaches to the Study of Crustacea*, 281–289.

Eder, E., Hödl, W., & Gottwald, R. (1997). Distribution and phenology of large branchiopods in Austria. *Hydrobiologia*, *359*, 13–22.

Eder, E., Schernhammer, T., Zuna-Kratky, T., & Nüsken, U. (2014). Temporäre Gewässer und ihre naturschutzfachliche Bedeutung. *Denisia*, *33*, 251–264.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294x.2005.02553.x

Evans, M. E. K., & Dennehy, J. J. (2005). GERM BANKING: BET-HEDGING AND VARIABLE RELEASE FROM EGG AND SEED DORMANCY. *The Quarterly Review of Biology*, *80*(4), 431–451. https://doi.org/10.1086/498282

Fong, J. J., Blom, M. P. K., Aowphol, A., McGuire, J. A., Sutcharit, C., & Soltis, P. S. (2023). Editorial: Recent advances in museomics: revolutionizing biodiversity research. *Frontiers in Ecology and Evolution*, *11*. https://doi.org/10.3389/fevo.2023.1188172

Francis, R. M. (2016). pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*(1), 27–32. https://doi.org/10.1111/1755-0998.12509

Gansauge, M., Aximu-Petri, A., Nagel, S., & Meyer, M. (2020). Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nature Protocols*, *15*(8), 2279–2300. https://doi.org/10.1038/s41596-020-0338-0

Gansauge, M., & Meyer, M. (2013). Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols*, *8*(4), 737–748. https://doi.org/10.1038/nprot.2013.038

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1207.3907

Gauthier, J., Pajkovic, M., Neuenschwander, S., Kaila, L., Schmid, S., Orlando, L., & Alvarez, N. (2020). Museomics identifies genetic erosion in two butterfly species across the 20th century in Finland. *Molecular Ecology Resources*, *20*(5), 1191–1205. https://doi.org/10.1111/1755-0998.13167

Glenn, T. C., Nilsen, R. A., Kieran, T. J., Sanders, J. G., Bayona-Vásquez, N. J., Finger, J. W., Pierson, T. W., Bentley, K. E., Hoffberg, S. L., Louha, S., Leon, F. J. G., Del Rio Portilla, M. A., Reed, K. D., Anderson, J. L., Meece, J. K., Aggrey, S. E., Rekaya, R., Alabady, M., Belanger, M., . . . Faircloth, B. C. (2019). Adapterama I: universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorically-indexed Illumina libraries (iTru & iNext). *PeerJ*, *7*, e7755. https://doi.org/10.7717/peerj.7755

Huson, D. H., & Bryant, D. (2005). Application of phylogenetic networks in evolutionary Studies. *Molecular Biology and Evolution*, *23*(2), 254–267. https://doi.org/10.1093/molbev/msj030

Jeffery, N. W. (2015). *Genome size diversity and evolution in the Crustacea* [PhD Thesis]. University of Guelph, Ontario, Canada.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Kapp, J. D., Green, R. E., & Shapiro, B. (2021). A fast and efficient single-stranded genomic library preparation method optimized for ancient DNA. *Journal of Heredity*, *112*(3), 241–249. https://doi.org/10.1093/jhered/esab012

Kassambara, A., & Mundt, F. (2016). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses* [Dataset]. https://doi.org/10.32614/cran.package.factoextra

Krueger, F., James, F., Ewels, P., Afyounian, E., Weinstein, M., Schuster-Boeckler, B., Hulselmans, G., & sclamons. (2023). FelixKrueger/TrimGalore: v0.6.10 - Add default decompression path (0.6.10). Zenodo. https://doi.org/10.5281/zenodo.7598955

Lalueza-Fox, C. (2022). Museomics. *Current Biology*, *32*(21), R1214–R1215. https://doi.org/10.1016/j.cub.2022.09.019

Lang, P. L. M., Weiß, C. L., Kersten, S., Latorre, S. M., Nagel, S., Nickel, B., Meyer, M., & Burbano, H. A. (2020). Hybridization ddRAD-sequencing for population genomics of nonmodel plants using highly degraded historical specimen DNA. *Molecular Ecology Resources*, *20*(5), 1228–1247. https://doi.org/10.1111/1755-0998.13168

Lavretsky, P., DaCosta, J. M., Sorenson, M. D., McCracken, K. G., & Peters, J. L. (2019). ddRAD-seq data reveal significant genome-wide population structure and divergent genomic regions that distinguish the mallard and close relatives in North America. *Molecular Ecology*, *28*(10), 2594–2609. https://doi.org/10.1111/mec.15091

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: ANRPackage for multivariate analysis. *Journal of Statistical Software*, *25*(1). https://doi.org/10.18637/jss.v025.i01

Li, H., & Durbin, R. (2006). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Linck, E. B., Hanna, Z. R., Sellas, A., & Dumbacher, J. P. (2017). Evaluating hybridization capture with RAD probes as a tool for museum genomics with historical bird specimens. *Ecology and Evolution*, *7*(13), 4755–4767. https://doi.org/10.1002/ece3.3065

LöFfler, H. (1993). Anostraca, Notostraca, Laevicaudata and Spinicaudata of Pannonian region and in its Austrian area. *Hydrobiologia*, *264*(3), 169–174. https://doi.org/10.1007/bf00007287

Longhurst, A. R. (1955a). A review of the Notostraca. *Bulletin of the British Museum (Natural History) Zoology*, *3*, 1–57. https://doi.org/10.5962/bhl.part.4119.

Longhurst, A. R. (1955b). The reproduction and cytology of the Notostraca (Crustacea, Phyllopoda). *Proc Zool Soc London, 125(3/4),* 671-680.

Luchetti, A., Forni, G., Martelossi, J., Savojardo, C., Martelli, P. L., Casadio, R., Skaist, A. M., Wheelan, S. J., & Mantovani, B. (2021). Comparative genomics of tadpole shrimps (Crustacea, Branchiopoda, Notostraca): Dynamic genome evolution against the backdrop of morphological stasis. *Genomics*, *113*(6), 4163–4172. https://doi.org/10.1016/j.ygeno.2021.11.001

Lukić, D., Waterkeyn, A., Rabet, N., Mioduchowska, M., Geudens, B., Vanschoenwinkel, B., Brendonck, L., & Pinceel, T. (2019). High genetic variation and phylogeographic relations among Palearctic fairy shrimp populations reflect persistence in multiple southern refugia during Pleistocene ice ages and postglacial colonisation. *Freshwater Biology*, *64*(11), 1896–1907. https://doi.org/10.1111/fwb.13380

Malinsky, M., Trucchi, E., Lawson, D. J., & Falush, D. (2018). RADpainter and fineRADstructure: Population Inference from RADseq Data. *Molecular Biology and Evolution*, *35*(5), 1284–1290. https://doi.org/10.1093/molbev/msy023

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, *38*(10), 4647–4654. https://doi.org/10.1093/molbev/msab199

Mantovani, B., Cesari, M., Luchetti, A., & Scanabissi, F. (2008). Mitochondrial and nuclear DNA variability in the living fossil *Triops cancriformis* (Bosc, 1801) (Crustacea, Branchiopoda, Notostraca). *Heredity*, *100*(5), 496–505. https://doi.org/10.1038/hdy.2008.3

Mantovani, B., Cesari, M., & Scanabissi, F. (2004). Molecular taxonomy and phylogeny of the 'living fossil' lineages *Triops* and *Lepidurus* (Branchiopoda: Notostraca). *Zoologica Scripta*, *33*(4), 367–374. https://doi.org/10.1111/j.0300-3256.2004.00155.x

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Masharing, N., Sodhi, M., Chanda, D., Singh, I., Vivek, P., Tiwari, M., Kumari, P., & Mukesh, M. (2023). ddRAD sequencing based genotyping of six indigenous dairy cattle breeds of India to infer existing genetic diversity and population structure. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-32418-6

Meirmans, P. *GenoDive*. (n.d.). https://www.patrickmeirmans.com/software/GenoDive.html.

Merta, L., Zavadil, V., Sychra, J. (2016). Atlas rozšíření velkých lupenonožců České republiky. Editor: Karel Chobot.

Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., & Gurevich, A. (2018). Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*, *34*(13), i142–i150. https://doi.org/10.1093/bioinformatics/bty266

Mioduchowska, M., Goldyn, B., Czyz, M. J., Namiotko, T., Namiotko, L., Kur, J., Sell, J. (2018). Notes on genetic uniformity in the fairy shrimp *Branchipus schaefferi* Fischer, 1834 (Branchiopoda, Anostraca) from Poland. *North-Western Journal Of Zoology 14* (1): 127-129.

O'Connor, B. D., & Van Der Auwera, G. (2020). *Genomics Analysis with Spark, Docker and Clouds*. O'Reilly Media.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADSEQ: an inexpensive method for de novo SNP discovery and genotyping in Model and Non-Model species. *PLoS ONE*, *7*(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Picard Tools - by Broad Institute. (n.d.). https://broadinstitute.github.io/picard/

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., & Korobeynikov, A. (2020). Using SPAdes de novo assembler. *Current Protocols in Bioinformatics*, *70*(1). https://doi.org/10.1002/cpbi.102

Raxworthy, C. J., & Smith, B. T. (2021). Mining museums for historical DNA: advances and challenges in museomics. *Trends in Ecology & Evolution*, *36*(11), 1049–1060. https://doi.org/10.1016/j.tree.2021.07.009

Rodríguez-Flores, P. C., Jiménez-Ruiz, Y., Forró, L., Vörös, J., & García-París, M. (2017). Non-congruent geographic patterns of genetic divergence across European species of *Branchinecta* (Anostraca: Branchinectidae). *Hydrobiologia*, *801*(1), 47–57. https://doi.org/10.1007/s10750-017-3266-4

RStudio Team. (2024). *RStudio: Integrated development environment for R (Version 2024.04.2+764, "Chocolate Cosmos")* [Computer software]. RStudio, PBC. https://www.rstudio.com/.

Sainz-Escudero, L., Vila, M., Perea, S., & García-París, M. (2023). Large effective size as determinant of population persistence in Anostraca (Crustacea: Branchiopoda). *Conservation Genetics*, *24*(6), 675–692. https://doi.org/10.1007/s10592-023-01534-1

Savojardo, C., Luchetti, A., Martelli, P. L., Casadio, R., & Mantovani, B. (2018). Draft genomes and genomic divergence of two *Lepidurus* tadpole shrimp species (Crustacea, Branchiopoda, Notostraca). *Molecular Ecology Resources*, *19*(1), 235–244. https://doi.org/10.1111/1755-0998.12952

Scanabissi, F., & Mondini, C. (2002). A survey of the reproductive biology in Italian branchiopods. *Hydrobiologia*, *486*(1), 263–272. https://doi.org/10.1023/a:1021371306687

Schwentner, M., & Lörz, A. (2020). Population genetics of cold-water coral associated Pleustidae (Crustacea, Amphipoda) reveals cryptic diversity and recent expansion off Iceland. *Marine Ecology*, *42*(1). https://doi.org/10.1111/maec.12625

Schwentner, M., & Richter, S. (2015). Stochastic effects associated with resting egg banks lead to genetically differentiated active populations in large branchiopods from temporary water bodies. *Hydrobiologia*, *760*(1), 239–253. https://doi.org/10.1007/s10750-015-2382-2

Suchan, T., Pitteloud, C., Gerasimova, N. S., Kostikova, A., Schmid, S., Arrigo, N., Pajkovic, M., Ronikier, M., & Alvarez, N. (2016). Hybridization capture using RAD probes (HyRAD), a new tool for performing genomic analyses on collection specimens. *PLoS ONE*, *11*(3), e0151651. https://doi.org/10.1371/journal.pone.0151651

Takahashi, T., Nagano, A. J., Kawaguchi, L., Onikura, N., Nakajima, J., Miyake, T., Suzuki, N., Kanoh, Y., Tsuruta, T., Tanimoto, T., Yasui, Y., Oshima, N., & Kawamura, K. (2020). A ddRAD-based population genetics and phylogenetics of an endangered freshwater fish from Japan. *Conservation Genetics*, *21*(4), 641–652. https://doi.org/10.1007/s10592-020-01275-5

Tin, M. M., Economo, E. P., & Mikheyev, A. S. (2014). Sequencing Degraded DNA from Non-Destructively Sampled Museum Specimens for RAD-Tagging and Low-Coverage Shotgun Phylogenetics. *PLoS ONE*, *9*(5), e96793. https://doi.org/10.1371/journal.pone.0096793

Vanschoenwinkel, B., Waterkeyn, A., Nhiwatiwa, T., Pinceel, T., Spooren, E., Geerts, A., Clegg, B., & Brendonck, L. (2011). Passive external transport of freshwater invertebrates by elephant and other mud-wallowing mammals in an African savannah habitat. *Freshwater Biology*, *56*(8), 1606–1619. https://doi.org/10.1111/j.1365-2427.2011.02600.x

Velonà, A., Luchetti, A., Scanabissi, F., & Mantovani, B. (2009). Genetic variability and reproductive modalities in European populations of *Triops cancriformis* (Crustacea, Branchiopoda, Notostraca). *Italian Journal of Zoology*, *76*(4), 366–375. https://doi.org/10.1080/11250000902785314

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*, *33*(14), 2202–2204. https://doi.org/10.1093/bioinformatics/btx153

Wall, A., Campo, D., & Wetzer, R. (2014). Genetic utility of natural history museum specimens: endangered fairy shrimp (Branchiopoda, Anostraca). *ZooKeys*, *457*, 1–14. https://doi.org/10.3897/zookeys.457.6822

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., & Van Den Brand, T. (2007). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics* [Dataset]. https://doi.org/10.32614/cran.package.ggplot2

Zofkova, M. (2006). Evolutionary dynamics in ephemeral pools: Inferences from genetic architecture of large branchiopoda [PhD Thesis]. School of Animal Biology, University of Western Australia, Perth, Australia.

VII. SUPPLEMENTARY INFORMATION

Collection Metadata Table (CM Tab.): List of all investigated *Lepidurus apus* (Linnaeus, 1758) individuals and their collection metadata.

Indv. ID	Country	Sampling location	Lat_DezGra	Lat_DezGra	Collection date	Preparation type	Collector
28389a	Austria	Rabensburg; kleine Sutte auf Feld	48.654987	16.9123404	20230510	ethanol	Schwentner, M.; Schernhammer, T.
28389b	Austria	Rabensburg; kleine Sutte auf Feld	48.654987	16.9123404	20230510	ethanol	Schwentner, M.; Schernhammer, T.
28389c	Austria	Rabensburg; kleine Sutte auf Feld	48.654987	16.9123404	20230510	ethanol	Schwentner, M.; Schernhammer, T.
28383	Austria	Naturreservat Marchauen; Storchenwiese	48.283	16.9055	20230510	ethanol	Schwentner, M.; Schernhammer, T.
28386	Austria	Pulverturm (hinter Weidentümpel)	48.27259586	16.915391	20230509	ethanol	Schwentner, M.; Schernhammer, T.
27644a	Austria	Niederösterreich, Marchegg, Pulverturm	48.170245	16.94951	19990420	ethanol	Eder, Erich
27679b	Austria	SO Weichselparz, gegegnüber Altarm, Baumgarten-Niederösterreich			19940515	ethanol	Eder Erich
27680b	Austria	Pulverturm, 'Riederlacke', Marchegg-Niederösterreich	48.273159	16.910577	19940506	ethanol	Eder Erich
27913b	Austria	NÖ, Marchegg, Triopossenke	48.239052	16.939367	20020417	ethanol	Eder, E.
27913a	Austria	NÖ, Marchegg, Triopossenke	48.239052	16.939367	20020417	ethanol	Eder, E.
29127b	Austria	Marchegg, Pulverturm	48.272948	16.914714	19920428	ethanol	Gruber, J.
27679a	Austria	SO Weichselparz, gegegnüber Altarm, Baumgarten-Niederösterreich			19940515	ethanol	Eder Erich
29127a	Austria	Marchegg, Pulverturm	48.272948	16.914714	19920428	ethanol	Gruber, J.
29704	Slovakia	Celáre (site 370), temporary wetland on a meadow	48.1258014	19.4935017	20230410	ethanol	Devánová, Alzbeta
29700c	Slovakia	Vysoká pri Morave (site 115), puddles on the road through the flooded meadow	48.3139631	16.9056803	20210326	ethanol	Devánová, Alzbeta
29703	Slovakia	Šahy (site 363), temporary wetland on arable land	48.0717644	18.9660931	20230410	ethanol	Devánová, Alzbeta
29700a	Slovakia	Vysoká pri Morave (site 115), puddles on the road through the flooded meadow	48.3139631	16.9056803	20210326	ethanol	Devánová, Alzbeta
29701a	Slovakia	Hrušov 1 (site 404), temporary wetland in the pasture	48.4266458	21.8704656	20230412	ethanol	Devánová, Alzbeta
29702c	Slovakia	Svinice (site 125), temporary wetland on arable land	48.4605736	21.8820894	20210502	ethanol	Devánová, Alzbeta
29705d	Slovakia	Malý Horeš (site 399), temporary wetland in the pasture	48.3905019	21.9646647	20230412	ethanol	Devánová, Alzbeta
29700e	Slovakia	Vysoká pri Morave (site 115), puddles on the road through the flooded meadow	48.3139631	16.9056803	20210326	ethanol	Devánová, Alzbeta
29701b	Slovakia	Hrušov 1 (site 404), temporary wetland in the pasture	48.4266458	21.8704656	20230412	ethanol	Devánová, Alzbeta
29700b	Slovakia	Vysoká pri Morave (site 115), puddles on the road through the flooded meadow	48.3139631	16.9056803	20210326	ethanol	Devánová, Alzbeta
29705a	Slovakia	Malý Horeš (site 399), temporary wetland in the pasture	48.3905019	21.9646647	20230412	ethanol	Devánová, Alzbeta
29705e	Slovakia	Malý Horeš (site 399), temporary wetland in the pasture	48.3905019	21.9646647	20230412	ethanol	Devánová, Alzbeta
29700d	Slovakia	Vysoká pri Morave (site 115), puddles on the road through the flooded meadow	48.3139631	16.9056803	20210326	ethanol	Devánová, Alzbeta

Indv. ID	Country	Sampling location	Lat_DezGra	Lat_DezGra	Collection date	Preparation type	Collector
29702f	Slovakia	Svinice (site 125), temporary wetland on arable land	48.4605736	21.8820894	20210502	ethanol	Devánová, Alzbeta
29705c	Slovakia	Malý Horeš (site 399), temporary wetland in the pasture	48.3905019	21.9646647	20230412	ethanol	Devánová, Alzbeta
29249b	Slovakia	Hrušov 2 (site 405), temporary wetland in the pasture under old oaks	48.4179025	21.8755133	20231005	ethanol	Devánová, Alzbeta
29702g	Slovakia	Svinice (site 125), temporary wetland on arable land	48.4605736	21.8820894	20210502	ethanol	Devánová, Alzbeta
29249a	Slovakia	Hrušov 2 (site 405), temporary wetland in the pasture under old oaks	48.4179025	21.8755133	20231005	ethanol	Devánová, Alzbeta
29705b	Slovakia	Malý Horeš (site 399), temporary wetland in the pasture	48.3905019	21.9646647	20230412	ethanol	Devánová, Alzbeta
29702d	Slovakia	Svinice (site 125), temporary wetland on arable land	48.4605736	21.8820894	20210502	ethanol	Devánová, Alzbeta
29702e	Slovakia	Svinice (site 125), temporary wetland on arable land	48.4605736	21.8820894	20210502	ethanol	Devánová, Alzbeta
29702b	Slovakia	Svinice (site 125), temporary wetland on arable land	48.4605736	21.8820894	20210502	ethanol	Devánová, Alzbeta
29702a	Slovakia	Svinice (site 125), temporary wetland on arable land	48.4605736	21.8820894	20210502	ethanol	Devánová, Alzbeta
30088	Austria				20130421	undenatured ethanol	Pass, T.
30089	Austria				20130421	undenatured ethanol	Pass, T.
FFPE 29127cc	Austria	Marchegg, Pulverturm	48.272948	16.914714	19920428	ethanol	Gruber, J.
FFPE 29127c	Austria	Marchegg, Pulverturm	48.272948	16.914714	19920428	ethanol	Gruber, J.
FFPE 27912a	Austria	NÖ, Baumgarten, Gassammelstation	48.32016	16.87211	20020410	ethanol	Eder, E.
FFPE 29127d	Austria	Marchegg, Pulverturm	48.272948	16.914714	19920428	ethanol	Gruber, J.
FFPE 27912b	Austria	NÖ, Baumgarten, Gassammelstation	48.32016	16.87211	20020410	ethanol	Eder, E.
FFPE 29127e	Austria	Marchegg, Pulverturm	48.272948	16.914714	19920428	ethanol	Gruber, J.
FFPE 27707b	Austria	Marchegg			19610405	ethanol (formerly in formol)	Vornatscher
FFPE 27708	Austria	Marchegg			19610421	ethanol (formerly in formol)	Vornatscher
29127cc	Austria	Marchegg, Pulverturm	48.272948	16.914714	19920428	ethanol	Gruber, J.
29127c	Austria	Marchegg, Pulverturm	48.272948	16.914714	19920428	ethanol	Gruber, J.
27912a	Austria	NÖ, Baumgarten, Gassammelstation	48.32016	16.87211	20020410	ethanol	Eder, E.
29127d	Austria	Marchegg, Pulverturm	48.272948	16.914714	19920428	ethanol	Gruber, J.
27912b	Austria	NÖ, Baumgarten, Gassammelstation	48.32016	16.87211	20020410	ethanol	Eder, E.
29127e	Austria	Marchegg, Pulverturm	48.272948	16.914714	19920428	ethanol	Gruber, J.
27707b	Austria	Marchegg			19610405	ethanol (formerly in formol)	Vornatscher
27708	Austria	Marchegg			19610421	ethanol (formerly in formol)	Vornatscher
26710b	Austria	Marchegg, Weide			19650428	ethanol	
27707a	Austria	Marchegg			19610405	ethanol (formerly in formol)	Vornatscher
26714i	Austria	Marchegg			19620504	ethanol	

Indv. ID	Country	Sampling location	Lat_DezGra	Lat_DezGra	Collection date	Preparation type	Collector
26719b	Austria	Marchegg, Pulverturm Leitung	48.272868	16.915282	19640424	ethanol (formerly in formol)	
27716a	Austria	Marchegg, Pulverturm-Leitung			19630508	ethanol (formerly in formol)	Vornatscher
27693b	Austria	Parndorf			19660413	ethanol (formerly in formol)	Vornatscher
27686b	Austria	Parndorf, Rav. 3, Baum			19660427	ethanol (formerly in formol)	Vornatscher
27709	Austria	Marchegg			19670816	ethanol (formerly in formol)	Vornatscher
27669d	Austria	Marchegg, Wiese bei Abzweigung Angern ('Cyzicus-Wiese')	48.278711	16.896738	19950419	ethanol	Eder Erich, Hödl Walter
27656b	Austria	Niederösterreich, Marchegg, Cyzikuswiese	48.278711	16.896738	19970417	ethanol	Eder, Erich
27704c	Austria	Marchegg			19800416	ethanol (formerly in formol)	Vornatscher
27662a	Austria	Niederösterreich, Marchegg, Cyzikuswiese	48.278711	16.896738	19950413	ethanol	Eder, Erich
29127f	Austria	Marchegg, Pulverturm	48.272948	16.914714	19920428	ethanol	Gruber, J.
27660b	Austria	Niederösterreich, Marchegg, Pulverturm	48.272948	16.914714	19950322	ethanol	Eder, Erich
29127dd	Austria	Marchegg, Pulverturm	48.272948	16.914714	19920428	ethanol	Gruber, J.
27661	Austria	Niederösterreich, Marchegg, Pulverturm	48.272948	16.914714	19950330	ethanol	Eder, Erich
27646	Austria	Burgenland, Parndorf, Design-Center	47.97611111	16.85416667	19990408	ethanol	Eder, Erich
26719a	Austria	Marchegg, Pulverturm Leitung	48.272868	16.915282	19640424	ethanol (formerly in formol)	
27716b	Austria	Marchegg, Pulverturm-Leitung			19630508	ethanol (formerly in formol)	Vornatscher
29127ee	Austria	Marchegg, Pulverturm	48.272948	16.914714	19920428	ethanol	Gruber, J.
26712b	Austria	Marchegg, Wegmulde			19630508	ethanol	
26711b	Austria	Marchegg, Wegmulde			19650510 (?)	ethanol	
26718b	Austria	Marchegg, Pulverturm	48.272868	16.915282	19630419	ethanol (formerly in formol)	
27677	Austria	Marchegg, Wiese nördl. Weg 9248 1a	48.283535	16.900314	19950325	ethanol	Eder Erich
27660a	Austria	Niederösterreich, Marchegg, Pulverturm	48.272948	16.914714	19950322	ethanol	Eder, Erich
27713b	Austria	Marchegg, Weide			19630517	ethanol (formerly in formol)	Vornatscher
27669c	Austria	Marchegg, Wiese bei Abzweigung Angern ('Cyzicus-Wiese')	48.278711	16.896738	19950419	ethanol	Eder Erich, Hödl Walter
27690a	Austria	Parndorf			19620425	ethanol (formerly in formol)	Vornatscher
27698c	Austria	Parndorf			19630421	ethanol (formerly in formol)	Vornatscher

Indv. ID	Country	Sampling location	Lat_DezGra	Lat_DezGra	Collection date	Preparation type	Collector
27645	Austria	Burgenland, Parndorf, Design-Center	47.97611111	16.85416667	19990422	ethanol	Eder, Erich

Laboratory Metadata Table (LM Tab.): List of all investigated *Lepidurus apus* (Linnaeus, 1758) individuals and their laboratory metadata.

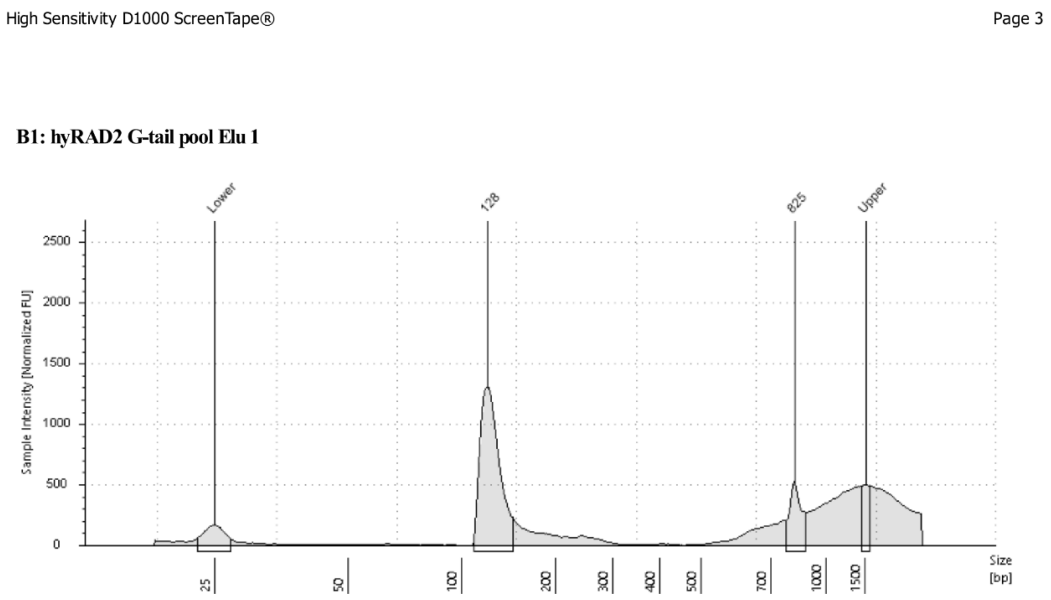
Indv. ID	Extraction tissue	Beads ratio	dsDNA [ng/µl]	ssDNA [ng/µl]	DNA [ng/µl]	Library prep. method	NGS library	Library ID	EcoRI adapter	MspI adapter	y-yoke adapter	GA 1st adapter	GA 2nd adapter	GA 1st adapter barcode	Primer rev	Primer fwd	DNA library [ng/µl]
28389a	3 legs	1x	80.4	-	-	ddRAD	ddRAD	RAD01	45	45					**B72**	**D508**	**7.5**
28389b	3 legs	1x	69.6	-	-	ddRAD	ddRAD	RAD01	42	42					B72	D508	7.5
28389c	3 legs	1x	87.6	-	-	ddRAD	ddRAD	RAD01	37	37					B72	D508	7.5
28383	3 legs	1x	44.4	-	-	ddRAD	ddRAD	RAD01	33	33					B72	D508	7.5
28386	3 legs	1x	22.6	-	-	ddRAD	ddRAD	RAD01	26	26					B72	D508	7.5
27644a	10 legs	1.5x	232	-	too high	ddRAD	ddRAD	RAD02	41	45					**B71**	**D507**	**0.298**
27679b	10 legs	1.5x	97.2	-	-	ddRAD	ddRAD	RAD02	43	42					B71	D507	0.298
27680b	10 legs	1.5x	91.2	-	-	ddRAD	ddRAD	RAD02	35	37					B71	D507	0.298
27913b	10 legs	1.5x	90.2	-	-	ddRAD	ddRAD	RAD02	32	33					B71	D507	0.298
27913a	10 legs	1.5x	68.4	-	-	ddRAD	ddRAD	RAD02	29	26					B71	D507	0.298
29127b	10 legs	1.5x	68	-	-	ddRAD	ddRAD	RAD02	25	23					B71	D507	0.298
27679a	10 legs	1.5x	66.2	-	-	ddRAD	ddRAD	RAD02	19	14					B71	D507	0.298
29127a	10 legs	1.5x	57.6	-	-	ddRAD	ddRAD	RAD02	12	8					B71	D507	0.298
29704	1 leg	1x	97	-	-	ddRAD	ddRAD	RAD11	45	8					**B68**	**D504**	**6.5**
29700c	3 legs	1x	69.4	-	-	ddRAD	ddRAD	RAD11	11	14					B68	D504	6.5
29703	1 leg	1x	62.6	-	-	ddRAD	ddRAD	RAD11	14	23					B68	D504	6.5
29700a	3 legs	1x	60.4	-	-	ddRAD	ddRAD	RAD11	23	26					B68	D504	6.5
29701a	1 leg	1x	58.8	-	-	ddRAD	ddRAD	RAD11	26	33					B68	D504	6.5
29702c	1 leg	1x	58.4	-	-	ddRAD	ddRAD	RAD11	33	37					B68	D504	6.5
29705d	1 leg	1x	56.2	-	-	ddRAD	ddRAD	RAD11	37	42					B68	D504	6.5
29700e	3 legs	1x	55.6	-	-	ddRAD	ddRAD	RAD11	42	45					B68	D504	6.5
29701b	1 leg	1x	50.2	-	-	ddRAD	ddRAD	RAD12	41	8					B67	D503	0.93
29700b	3 legs	1x	50	-	-	ddRAD	ddRAD	RAD12	43	14					B67	D503	0.93
29705a	1 leg	1x	48.6	-	-	ddRAD	ddRAD	RAD12	35	23					B67	D503	0.93
29705e	1 leg	1x	45.8	-	-	ddRAD	ddRAD	RAD12	32	26					B67	D503	0.93
29700d	3 legs	1x	45.6	-	-	ddRAD	ddRAD	RAD12	29	33					B67	D503	0.93
29702f	1 leg	1x	45.2	-	-	ddRAD	ddRAD	RAD12	25	37					B67	D503	0.93
29705c	1 leg	1x	44.8	-	-	ddRAD	ddRAD	RAD12	19	42					B67	D503	0.93
29249b	1 leg	1x	44.2	-	-	ddRAD	ddRAD	RAD12	12	45					B67	D503	0.93
29702g	1 leg	1x	34.2	-	-	ddRAD	ddRAD	RAD13	45	45					**B66**	**D502**	**0.652**

Indv. ID	Extraction tissue	Beads ratio	dsDNA [ng/µl]	ssDNA [ng/µl]	DNA [ng/µl]	Library prep. method	NGS library	Library ID	EcoRI adapter	MspI adapter	y-yoke adapter	GA 1st adapter	GA 2nd adapter	GA 1st adapter barcode	Primer rev	Primer fwd	DNA library [ng/µl]
29249a	1 leg	1x	32.2	-	-	ddRAD	ddRAD	RAD13	42	42					B66	D502	0.652
29705b	1 leg	1x	28.2	-	-	ddRAD	ddRAD	RAD13	37	37					B66	D502	0.652
29702d	1 leg	1x	27.4	-	-	ddRAD	ddRAD	RAD13	33	33					B66	D502	0.652
29702e	1 leg	1x	22.2	-	-	ddRAD	ddRAD	RAD13	26	26					B66	D502	0.652
29702b	1 leg	1x	20.2	-	-	ddRAD	ddRAD	RAD13	23	23					B66	D502	0.652
29702a	1 leg	1x	17.3	-	-	ddRAD	ddRAD	RAD13	14	14					B66	D502	0.652
30088	3 legs	1x	71	-	-	ddRAD	ddRAD	RAD15	41	37					B66	D502	0.652
30089	3 legs	1x	120	-	-	ddRAD	ddRAD	RAD15	43	33					B66	D502	0.652
FFPE_29127cc	6-7 legs	1.5x	7.5	49.84	56.2	FFPE+G-tailing+Ultra II	hyRAD	Gt02			1_01				B65	D508	**21**
FFPE_29127c	10 legs	1.5x	16.6	37.6	54.2	FFPE+G-tailing+Ultra II	hyRAD	Gt02			1_02				B66	D507	21
FFPE 27912a	10 legs	1.5x	57.6	-	too high	FFPE+G-tailing+Ultra II	hyRAD	Gt02			1_03				B67	D506	21
FFPE_29127d	10 legs	1.5x	8.28	30.52	38.8	FFPE+G-tailing+Ultra II	hyRAD	Gt02			1_04				B68	D505	21
FFPE_27912b	10 legs	1.5x	68	-	too high	FFPE+G-tailing+Ultra II	hyRAD	Gt02			1_05				B69	D504	21
FFPE_29127e	10 legs	1.5x	5.78	15.42	21.2	FFPE+G-tailing+Ultra II	hyRAD	Gt02			1_06				B70	D503	21
FFPE_27707b	10 legs	2x	2.1	7.48	9.58	FFPE+G-tailing+Ultra II	hyRAD	Gt02			1_07				B71	D502	21
FFPE_27708	10 legs	2x	1.81	6.81	8.62	FFPE+G-tailing+Ultra II	hyRAD	Gt02			1_08				B72	D501	21
29127cc	6-7 legs	1.5x	7.5	49.84	56.2	G-tailing+Ultra II	hyRAD	Gt02			1_08				B72	D508	21
29127c	10 legs	1.5x	16.6	37.6	54.2	G-tailing+Ultra II	hyRAD	Gt02			1_07				B71	D507	21
27912a	10 legs	1.5x	57.6	-	too high	G-tailing+Ultra II	hyRAD	Gt02			1_06				B70	D506	21
29127d	10 legs	1.5x	8.28	30.52	38.8	G-tailing+Ultra II	hyRAD	Gt02			1_05				B69	D505	21
27912b	10 legs	1.5x	68	-	too high	G-tailing+Ultra II	hyRAD	Gt02			1_04				B68	D504	21
29127e	10 legs	1.5x	5.78	15.42	21.2	G-tailing+Ultra II	hyRAD	Gt02			1_03				B67	D503	21
27707b	10 legs	2x	2.1	7.48	9.58	G-tailing+Ultra II	hyRAD	Gt02			1_02				B66	D502	21
27708	10 legs	2x	1.81	6.81	8.62	G-tailing+Ultra II	hyRAD	Gt02			1_01				B65	D501	21
26710b	10 legs	2x	1.28	8	9.28	G-tailing+FS	hyRAD	Gt03			1_01				B72	D507	**50**
27707a	10 legs	2x	4.82	15.38	20.2	G-tailing+FS	hyRAD	Gt03			1_02				B71	D505	50
26714i	6-7 legs	2x	2.79	8.32	10.9	G-tailing+FS	hyRAD	Gt03			1_03				B70	D503	50
26719b	10 legs	2x	1.29	4.25	5.54	G-tailing+FS	hyRAD	Gt03			1_04				B69	D501	50

Indv. ID	Extraction tissue	Beads ratio	dsDNA [ng/µl]	ssDNA [ng/µl]	DNA [ng/µl]	Library prep. method	NGS library	Library ID	EcoRI adapter	MspI adapter	y-yoke adapter	GA 1st adapter	GA 2nd adapter	GA 1st adapter barcode	Primer rev	Primer fwd	DNA library [ng/µl]
27716a	10 legs	2x	0.27	1.83	2.1	G-tailing+FS	hyRAD	Gt03			1_05				B68	D508	50
27693b	10 legs	2x	2.94	9.56	12.5	G-tailing+FS	hyRAD	Gt03			1_06				B67	D506	50
27686b	10 legs	2x	6.74	35.66	42.4	G-tailing+FS	hyRAD	Gt03			1_07				B66	D504	50
27709	10 legs	2x	5.02	18.78	23.8	G-tailing+FS	hyRAD	Gt03			1_08				B65	D502	50
27669d	6-7 legs	1.5x	1.53	31.27	32.8	G-tailing+FS	hyRAD	Gt03			1_01				B72	D502	50
27656b	6-7 legs	1.5x	2.12	9.18	11.3	G-tailing+FS	hyRAD	Gt03			1_02				B71	D504	50
27704c	10 legs	1.5x	1.43	8.23	9.66	G-tailing+FS	hyRAD	Gt03			1_03				B70	D506	50
27662a	6-7 legs	1.5x	0.255	3.745	4	G-tailing+FS	hyRAD	Gt03			1_04				B69	D508	50
29127f	10 legs	1.5x	10.1	60.1	70.2	G-tailing+FS	hyRAD	Gt03			1_05				B68	D501	50
27660b	10 legs	1.5x	1.91	15.19	17.1	G-tailing+FS	hyRAD	Gt03			1_06				B67	D503	50
29127dd	6-7 legs	1.5x	7.2	-	too high	G-tailing+FS	hyRAD	Gt03			1_07				B66	D505	50
27661	10 legs	1.5x	1.14	4.84	5.98	G-tailing+FS	hyRAD	Gt03			1_08				B65	D507	50
27646	6-7 legs	1.5x	8.3	55.3	63.6	Gansauge & Meyer (2013)	hyRAD	GA1WDH				TL181/TL159	CL35/TL178		B72	D507	**51**
26719a	10 legs	2x	1.29	4.25	5.54	Gansauge & Meyer (2013)	hyRAD	GA1WDH				TL181/TL159	CL35/TL178		B66	D503	51
27716b	10 legs	2x	0.71	3.25	3.96	Gansauge & Meyer (2013)	hyRAD	GA1WDH				TL181/TL159	CL35/TL178		B68	D501	51
29127ee	6-7 legs	1.5x	6.3	32.32	38.8	Gansauge & Meyer (2013)	hyRAD	GA1WDH				TL181/TL159	CL35/TL178		B65	D502	51
26712b	10 legs	2x	0.374	3.786	4.16	Gansauge & Meyer (2013)	hyRAD	GA1WDH				TL181/TL159	CL35/TL178		B68	D503	51
26711b	10 legs	2x	1.05	4.17	5.22	Gansauge & Meyer (2013)	hyRAD	GA1WDH				TL181/TL159	CL35/TL178		B71	D508	51
26718b	10 legs	2x	1.01	4.09	5.1	Gansauge & Meyer (2013)	hyRAD	GA1WDH				TL181/TL159	CL35/TL178		B67	D504	51
27677	6-7 legs	1.5x	2.27	3.47	5.74	Gansauge & Meyer (2013)	hyRAD	GA1WDH				TL181/TL159	CL35/TL178		B67	D502	51
27660a	10 legs	1.5x	5.5	27.3	32.8	Gansauge & Meyer (2013) with barcode	hyRAD	GA1BWDH						TL181_01/TL159_01	B72	D504	**45**
27686b	10 legs	2x	6.74	35.66	42.4	Gansauge & Meyer (2013) with barcode	hyRAD	GA1BWDH						TL181_01/TL159_01	B68	D502	45
27644a	10 legs	1.5x	232	-	too high	Gansauge & Meyer (2013) with barcode	hyRAD	GA1BWDH						TL181_01/TL159_01	B69	D501	45

Indv. ID	Extraction tissue	Beads ratio	dsDNA [ng/µl]	ssDNA [ng/µl]	DNA [ng/µl]	Library prep. method	NGS library	Library ID	EcoRI adapter	MspI adapter	y-yoke adapter	GA 1st adapter	GA 2nd adapter	GA 1st adapter barcode	Primer rev	Primer fwd	DNA library [ng/µl]
27713b	10 legs	2x	5.48	20.72	26.2	Gansauge & Meyer (2013) with barcode	hyRAD	GA1BWDH						TL181_01/TL159_01	B70	D506	45
27669c	6-7 legs	1.5x	1.06	34.14	35.2	Gansauge & Meyer (2013) with barcode	hyRAD	GA1BWDH						TL181_01/TL159_01	B72	D508	45
27690a	10 legs	2x	2.24	9.56	11.8	Gansauge & Meyer (2013) with barcode	hyRAD	GA1BWDH						TL181_01/TL159_01	B69	D507	45
27698c	10 legs	2x	0.762	3.578	4.34	Gansauge & Meyer (2013) with barcode	hyRAD	GA1BWDH						TL181_01/TL159_01	B65	D508	45
27645	6-7 legs	1.5x	21	199	220	Gansauge & Meyer (2013) with barcode	hyRAD	GA1BWDH						TL181_01/TL159_01	B71	D507	45
29704	1 leg	1x	97	-	-	ddRAD+Biotinylation (Lang et al., 2020)	Baits	RAD10	15	14					B69	D505	
29700c	3 legs	1x	69.4	-	-	ddRAD+Biotinylation (Lang et al., 2020)	Baits	RAD10	15	14					B69	D505	
29703	1 leg	1x	62.6	-	-	ddRAD+Biotinylation (Lang et al., 2020)	Baits	RAD10	15	14					B69	D505	
29700a	3 legs	1x	60.4	-	-	ddRAD+Biotinylation (Lang et al., 2020)	Baits	RAD10	15	14					B69	D505	
29702c	1 leg	1x	58.4	-	-	ddRAD+Biotinylation (Lang et al., 2020)	Baits	RAD10	15	14					B69	D505	
29705c	1 leg	1x	44.8	-	-	ddRAD+Biotinylation (Lang et al., 2020)	Baits	RAD10	15	14					B69	D505	
28383	3 legs	1x	44.4	-	-	ddRAD+Biotinylation (Lang et al., 2020)	Baits	RAD10	15	14					B69	D505	
28386	3 legs	1x	22.6	-	-	ddRAD+Biotinylation (Lang et al., 2020)	Baits	RAD10	15	14					B69	D505	
28389b	3 legs	1x	69.6	-	-		WGSeq	WGS1									

Supplementary Figure 1: High Sensitivity D1000 DNA fragment measurement of library Gt02 on a 4200 TapeStation System by Agilent.



Supplementary Table S1: The protocols of the homemade washing buffers for the hybridisation capture / hyRAD protocol after Crates et al. (2020) and Lang et al. (2020).

TEN buffer	**1x [µl]**
NaCl (5 M)	200
Tris-HCl (1 M)	10
EDTA (0.5 mM)	2
HPLC	788
1x SSC/0.1% SDS	**1x [µl]**
SSC	50
SDS 20%	5
HPLC	945
0.5x SSC/0.1% SDS	**1x [µl]**
SSC	25
SDS 20%	5
HPLC	970
0.1x SSC/0.1% SDS	**1x [µl]**
SSC	5
SDS 20%	5
HPLC	990
Melting solution	**1x [µl]**
NaOH (1 M)	125
Tween 2%	25
HPLC	850
TT buffer	**1x [µl]**
Tris-HCl (1 M)	1
10% Tween-20	1

HPLC	998

Supplementary Table S2: Parameters for the assembly tool *ipyrad* (Eaton and Overcast, 2020).

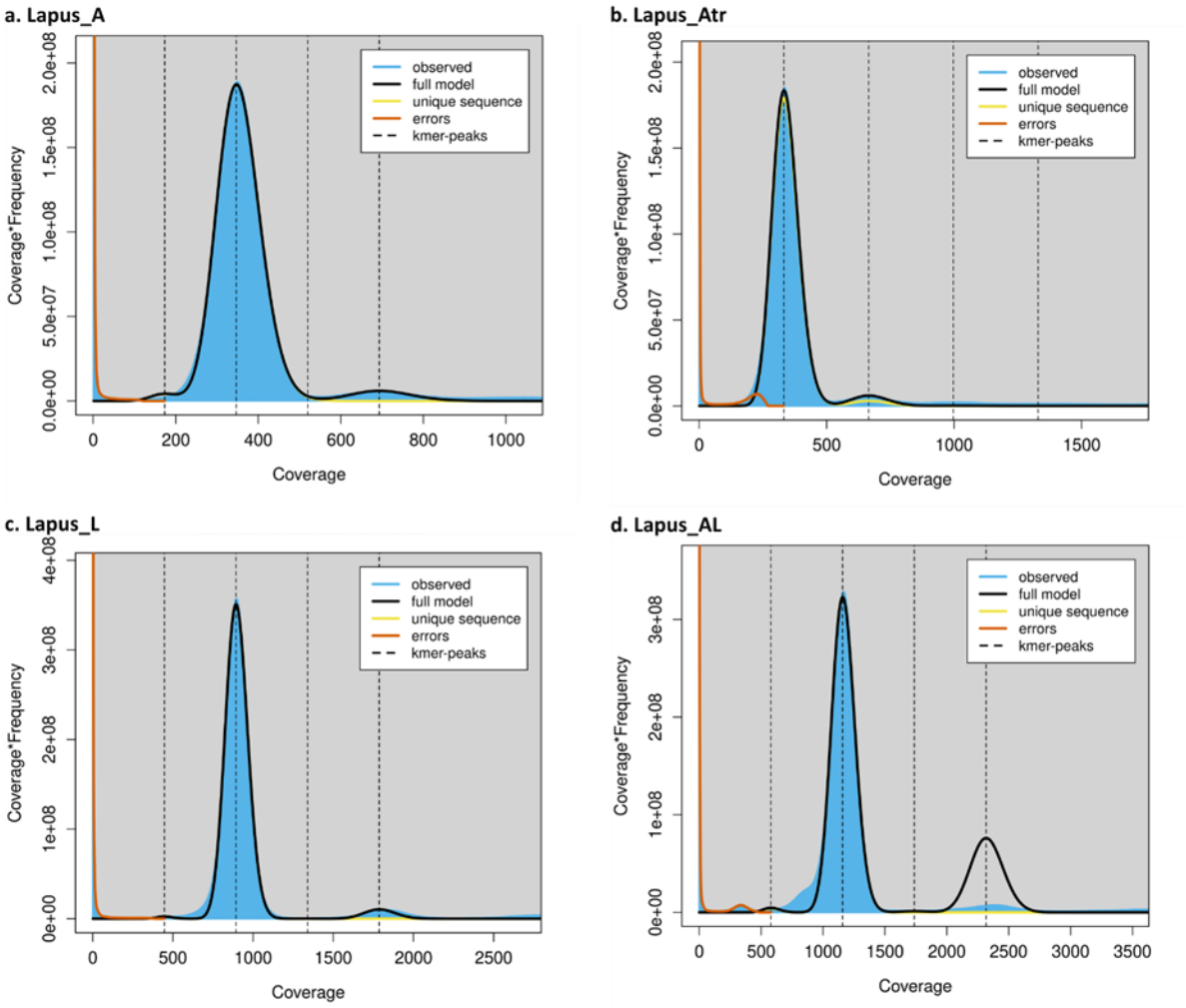
------- ipyrad params file (v.0.9.95) ----------	
ipyrad_Lapus_to11	[0] [assembly_name]: Assembly name. Used to name output directories for assembly steps
/media/inter/aloewenstein/ddRAD_analysis/ipyrad/tryout11	[1] [project_dir]: Project dir (made in curdir if not present)
	[2] [raw_fastq_path]: Location of raw non-demultiplexed fastq files
	[3] [barcodes_path]: Location of barcodes file
/media/inter/aloewenstein/ddRAD_analysis/ ddrad_samples_final/*.fastq	[4] [sorted_fastq_path]: Location of demultiplexed/sorted fastq files
reference	[5] [assembly_method]: Assembly method (denovo, reference)
/media/inter/aloewenstein/Lapus_assembly_Augustina_trimmed /output/Lapus_Atr_EuNH_exclHs_larger500bp.fa	[6] [reference_sequence]: Location of reference sequence file
pairddrad	[7] [datatype]: Datatype (see docs): rad, gbs, ddrad, etc.
AATT, CG	[8] [restriction_overhang]: Restriction overhang (cut1,) or (cut1, cut2)
5	[9] [max_low_qual_bases]: Max low quality base calls (Q<20) in a read
33	[10] [phred_Qscore_offset]: phred Q score offset (33 is default and very standard)
6	[11] [mindepth_statistical]: Min depth for statistical base calling
6	[12] [mindepth_majrule]: Min depth for majority-rule base calling
10000	[13] [maxdepth]: Max cluster depth within samples
0.85	[14] [clust_threshold]: Clustering threshold for de novo assembly
0	[15] [max_barcode_mismatch]: Max number of allowable mismatches in barcodes
2	[16] [filter_adapters]: Filter for adapters/primers (1 or 2=stricter)
35	[17] [filter_min_trim_len]: Min length of reads after adapter trim
2	[18] [max_alleles_consens]: Max alleles per site in consensus sequences
0.05	[19] [max_Ns_consens]: Max N's (uncalled bases) in consensus
0.05	[20] [max_Hs_consens]: Max Hs (heterozygotes) in consensus
17	[21] [min_samples_locus]: Min samples per locus for output
0.1	[22] [max_SNPs_locus]: Max SNPs per locus
8	[23] [max_Indels_locus]: Max of indels per locus
0.5	[24] [max_shared_Hs_locus]: Max heterozygous sites per locus
0, 0, 0, 0	[25] [trim_reads]: Trim raw read edges (R1>, <R1, R2>, <R2) (see docs)
0, 0, 0, 0	[26] [trim_loci]: Trim locus edges (see docs) (R1>, <R1, R2>, <R2)
*	[27] [output_formats]: Output formats (see docs)
	[28] [pop_assign_file]: Path to population assignment file
	[29] [reference_as_filter]: Reads mapped to this reference are removed in step 3

Supplementary Table S3: Tools and software used for the mapping script of the hyRAD assembly.

Name of tool	**Version**	**Authors**	**Command lines**
bwa	0.7.13	Li and Durbin (2009)	bwa aln bwa sampe
samtools	1.18	Danecek et al. (2021)	samtools view samtools sort samtools index samtools flagstat

Name of tool	Version	Authors	Command lines
			samtools merge
picard	2.27.5	http://broadinstitute.github.io/picard	AddOrReplaceReadGroups
GATK	3.8.1	O'Connor and Van Der Auwera (2020)	RealignerTargetCreator IndelRealigner
freebayes	3.0.0	Garrison and Marth (2012)	freebayes -f -F 0.3
vcftools	0.1.13	Danecek et al. (2011)	--min-alleles 2 –max-alleles 2 –remove-indels –minDP –max-missing –recode --out

Supplementary Figure 2: *Genomescope* plots of final assemblies showing the transformed genome coverage per frequency.



Supplementary Table S4: *QUAST* and *BUSCO* output for the raw and post filtering, final assemblies (excluding < 500 bp contigs and off-target DNA).

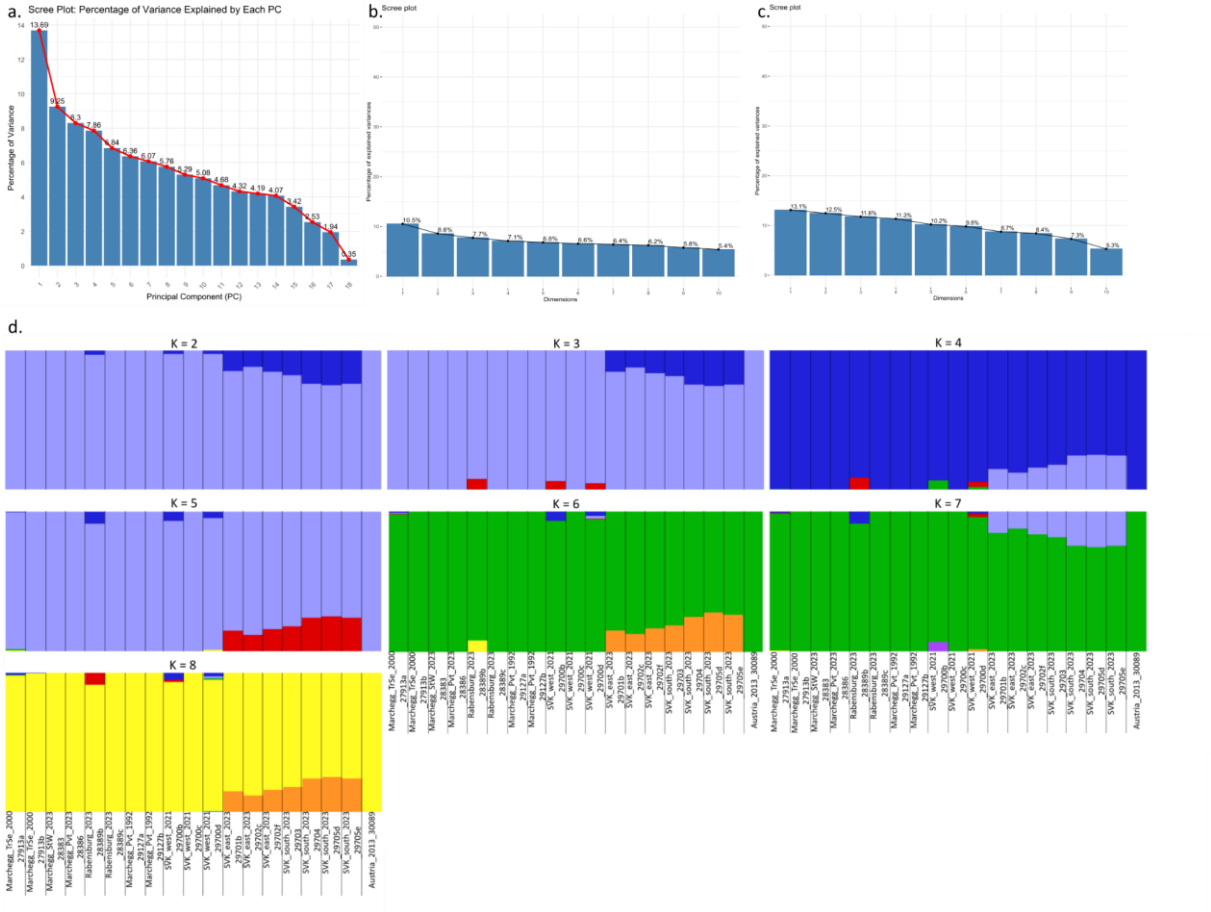
	Raw Lapus_A assembly	Final Lapus_A assembly	Raw Lapus_Atr assembly	Final Lapus_Atr assembly	Raw Lapus_L assembly	Final Lapus_L assembly	Raw Lapus_AL assembly	Final Lapus_AL assembly
QUAST								
# contigs (>= 0 bp)	198445	17203	191316	16810	237824	22461	361020	32408
# contigs (>= 1000 bp)	8840	6185	8600	6041	18261	10477	22297	13107
# contigs (>= 5000 bp)	1885	1815	1876	1828	4933	3473	5288	3749
# contigs (>= 10000 bp)	1207	1206	1221	1221	2967	2278	3002	2320
# contigs (>= 25000 bp)	661	660	668	668	1216	1014	1196	1003
# contigs (>= 50000 bp)	388	387	390	390	519	453	497	439
Total length (>= 0 bp)	132453078	84622163	130660473	84304952	210405185	122307307	248459871	128799939
Total length (>= 1000 bp)	82318230	77278442	81781670	77136612	149710587	114006967	153746813	115667576
Total length (>= 5000 bp)	69649506	69103211	69491326	69195940	123580586	100336587	121070654	97871834
Total length (>= 10000 bp)	64961327	64856548	64948129	64948129	109665202	91832552	104980870	87734226
Total length (>= 25000 bp)	56116888	56012109	56001149	56001149	82401502	71831207	76752165	66884121
Total length (>= 50000 bp)	46464407	46359628	46228176	46228176	58514845	52561332	52881398	47729331
# contigs	24033	17203	23507	16810	38645	22461	52662	32408
Largest contig	778048	778048	604501	604501	551951	551951	458499	458499
Total length	92585689	84622163	91848109	84304952	163894043	122307307	174591209	128799939
GC (%)	42.88	42.10	42.82	42.06	43.64	40.34	43.73	40.58
N50	50930	63078	51150	63059	25419	36433	18404	26845
N90	923	1171	927	1208	1120	1511	887	984
L50	385	316	384	318	1198	657	1690	907
L90	9893	5146	9517	4887	16173	7226	25895	13362
# N's per 100 kb	31.46	28.45	30.86	27.65	236.23	188.96	223.54	211.35
BUSCO								
Comp. %	97.4	97.3	97.4	97.3	97.7	97.7	97.3	97.3
Dupl. %	0.4	0.4	0.5	0.4	0.3	0.3	0.8	0.8
Frag. %	1.8	1.8	1.8	1.8	1.5	1.5	1.8	1.8
Missing %	0.8	0.9	0.8	0.9	0.8	0.8	0.9	0.9
Complete BUSCOs (C)	987	986	987	986	990	990	986	986

	Raw Lapus_A assembly	Final Lapus_A assembly	Raw Lapus_Atr assembly	Final Lapus_Atr assembly	Raw Lapus_L assembly	Final Lapus_L assembly	Raw Lapus_AL assembly	Final Lapus_AL assembly
Complete and single-copy BUSCOs (S)	983	982	982	982	987	987	978	978
Complete and duplicated BUSCOs (D)	4	4	5	4	3	3	8	8
Fragmented BUSCOs (F)	18	18	18	18	15	15	18	18
Missing BUSCOs (M)	8	9	8	9	8	8	9	9
Total BUSCO groups searched	1013	1013	1013	1013	1013	1013	1013	1013
Number of scaffolds	177384	17203	170669	17203	176338	22461	287021	32408
Number of contigs	178447	18246	171696	18246	181805	27926	293966	39308
Total length	120356631	84622163	118960634	84622163	155525371	122307307	187229512	128799939
Percent gaps	0.021%	0.028%	0.020%	0.02%	0.149%	0.19%	0.146%	0.211%
Scaffold N50	17 KB	63 KB	18 KB	63 KB	19 KB	36 KB	7 KB	26 KB
Contigs N50	16 KB	58 KB	17 KB	58 KB	15 KB	30 KB	5 KB	18 KB

Supplementary Table S5: Results of different *ipyrad* parameters for the ddRAD assembly.
[21] [min_samples_locus]: Min. samples per locus for output.
[22] [max_SNPs_locus]: Max. SNPs per locus.

	[21] 80% [22] 20%	[21] 90% [22] 20%	[21] 80% [22] 10%
Nbr. of loci	6850	5302	6845
Nbr. of SNPs	15150	11777	14900

Supplementary Figure 3: Scree plots for ddRAD sub-assembly PCAs and bar plots of admixture proportions of the ddRAD-seq 5163 SNPs. a. Final ddRAD assembly, 5163 SNPs. **b.** ddRAD assembly, excluding 289 possibly falsely called historic SNPs. **c.** ddRAD assembly, excluding 289 possibly falsely called historic SNPs and eastern/southern Slovakia individuals. **d.** Bar plots of admixture for K = 2-8.



Supplementary Figure 4: Scree plots for hyRAD sub-assembly PCAs.