



Dec 17, 2024

Single-tube DNA library preparations for museum butterfly specimens

DOI

<https://dx.doi.org/10.17504/protocols.io.yxmvm9k36l3p/v1>



Berenice Lafon¹, Eliette Reboud¹, Marie-Ka Tilak¹, Fabien Condamine¹

¹CNRS, Institut des Sciences de l'Evolution de Montpellier, Place Eugène Bataillon, Montpellier, France



Berenice J. Lafon

CNRS

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.yxmvm9k36l3p/v1>

Protocol Citation: Berenice Lafon, Eliette Reboud, Marie-Ka Tilak, Fabien Condamine 2024. Single-tube DNA library preparations for museum butterfly specimens. **protocols.io** <https://dx.doi.org/10.17504/protocols.io.yxmvm9k36l3p/v1>

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: November 20, 2024

Last Modified: December 17, 2024

Protocol Integer ID: 112484

Keywords: Historical DNA, Insect DNA, DNA purification, Whole genome sequencing, tube dna library preparations for museum butterfly, tube dna library preparation, house library preparation for the same dna, stranded dna fragment, dna from museum butterfly, short dna fragment, dna fragment, fragmented dna molecule, degraded dna, historical dna, tube library preparation, dna, quality dna, thermal inactivation of the enzyme, purification step, quality of genome assembly, genome assembly, tube protocol, column purification with thermal inactivation, tube protocol of carøe et al, same dna, purification, paleogenomic, replacing column purification, museum butterfly, enzyme, genome, assembly statistics for the genome, stranded dna, sequencing, specimen

Funders Acknowledgements:

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (project GAIA, agreement no. 851188).

Grant ID: 851188

Abstract

Protocols for single-tube library preparation can eliminate inter-reaction purifications, which are made possible by replacing column purification with thermal inactivation of enzymes (Zheng *et al.* 2011; Neiman *et al.* 2012). Such protocols allow libraries to be prepared much more quickly, with less manual handling, and reduce the economic cost of using modern, high-quality DNA. These approaches can be used for degraded DNA, where sequencing is difficult due to the limitations of handling chemically damaged and highly fragmented DNA molecules. The advantage of this method is that very short DNA fragments (less than 70 bp) are not lost during the purification steps. It also avoids the denaturation of ultrashort double-stranded DNA fragments (e.g. 25 bp), which can denature at 72°C depending on the sequence composition (Owczarzy *et al.* 1997), since the thermal inactivation of the enzymes is performed at 65°C instead of the usual 72°C (Zheng *et al.* 2011; Neiman *et al.* 2012). Here, we propose an adaptation to the single-tube protocol of Carøe *et al.* (2018), in which we use the adapters of Meyer & Kircher (2010) on historical (degraded) DNA from museum butterflies. We successfully extended the protocol from vertebrates to insects, and found that the Meyer & Kircher's adapters are more efficient in this protocol, resulting in fewer adapter dimers. This method allows the retention of very small fragments (<70 bp), which is of great interest in the growing field of paleogenomics and other applications studying degraded and historical DNA. By comparing the quality of genome assemblies with a classical in-house library preparation for the same DNA, we found similar (if not better) assembly statistics for the genomes assembled with the single-tube protocol.

Image Attribution

Pictured is a box of specimens of the genus *Luehdorfia japonica* from the Muséum National d'Histoire Naturelle (Paris, France). Photographed by Fabien L. Condamine.

Guidelines

Table: Adapters mix primers (Meyer & Kircher 2010)

A	B
Names	Sequences
IS1_adapter P5.F	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCG*A*T*C*T
IS2_adapter P7.F	G*T*G*A*CTGGAGTTCAGACGTGTGCTCTTCCG*A*T*C*T
IS3_adapter P5+P7.R	A*G*A*T*CGGAA*G*A*G*C
IS4_ind PCR P5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT
I1_ind PCR P7	CAAGCAGAAGACGGCATACGAGATcctgcgaGTGACTGGAGTTCAGACGTGT

* PTO = phosphothioate oligonucleotide bond

Recipes:

A
Oligo hybridization buffer (10X)
500 mM NaCl
10 mM Tris-Cl, pH 8.0
1 mM EDTA, pH 8.0

Adapter mix ready to use:

A	B	C	D
Reagent	Volume (μL)	[Final]	[Combined]
IS1_adapter_P7.F (500 μM)	20	200 μM	100 uM each
IS3_adapter_P5+P7.R (500 μM)	20	200 μM	100 uM each
Oligo hybridization buffer (10X)	5	1X	1X
H2O	5	-	-

Mix and incubate each reaction in a thermal cycler for 10 s at 95°C, followed by ramp from 95°C to 12°C decreasing of -0.1°C/s. Pool both reactions to obtain adapter mix ready to use.



Materials

Reagents:

- DNeasy Blood and Tissue Kit (Qiagen)
- Agencourt AMPure XP beads (Beckman Coulter, France SAS)
- Taq Polymerase (NEB cat#M0273S, 5 U/μL)
- 25 mM dNTPs (New England Biolabs, Ipswich, MA, USA)
- T4 DNA Polymerase (3 U/μL, New England Biolabs, cat#M0203S)
- T4 Polynucleotide Kinase (10 U/μL, New England Biolabs, cat#M0201S)
- 10X T4 DNA Ligase Reaction Buffer (New England Biolabs, cat#B0202S)
- Polyethyleneglycol 4000 (50%, Thermo Scientific Chemicals ref: A16151.30)
- T4 DNA Ligase (5.97 U/μL, New England Biolabs, cat#M0202S)
- 2X KAPA HiFi HotStart ReadyMix (Roche)
- Isothermal Amplification Buffer (New England Biolabs, cat#B0537S)
- Bst 2.0 Warmstart Polymerase (New England Biolabs, cat#M0538S)
- 10X Oligo Hybridation Buffer

Equipments:

- Centrifuge
- 0.5 μL LoBind tube
- Incubator (Eppendorf ThermoMixer® F1.5)
- Qubit 4.0 Fluorometer (Thermo Fischer Scientific)

Aim of the study

- 1 Today, advances in next-generation sequencing and library construction techniques provide a better approach to the study of historical (degraded) DNA (hereafter named: hDNA), which is often degraded and fragmented. In our laboratory, we rely on a well-tested protocol for library construction developed by Tilak *et al.* (2015), which was modified and adapted from the classical protocol of Meyer & Kircher (2010). This classical in-house protocol has been successfully performed for insect species for short read sequencing with Illumina, in particular for butterflies (e.g. Reboud *et al.* 2024). However, this protocol does not always work optimally with hDNA from museum specimens. Interestingly, new protocols have emerged, in particular the "single tube" method (e.g. Carøe *et al.* 2018). For hDNA, such a protocol is thought to be faster and less costly, and also reduces the loss of small fragments by eliminating several purification steps. Originally tested on vertebrate hDNA, we here test and propose an adaptation of the "single tube" protocol of Carøe *et al.* (2018) on hDNA from museum butterflies.

DNA extraction

1d

- 2 The specimen is a *Papilio machaon* (FC3013), a common swallowtail butterfly species distributed across the Palearctic. Total genomic DNA has been extracted from a leg using the *DNeasy Blood and Tissue Kit* from Qiagen. We followed the manufacturer's instructions.

1d

DNA was quantified using the Qubit 4.0 fluorometer (Thermo Fisher Scientific). DNA quality was also checked by running a 1% agarose gel or Agilent 2200 TapeStation System.

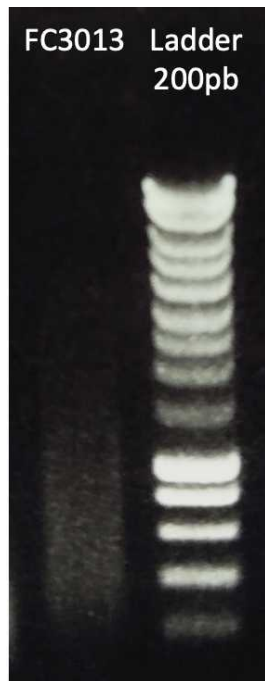


Figure 1. Pictured is an agarose gel showing the DNA of *Papilio machaon* (FC3013) after the Qiagen extraction. The DNA profile looks fragments with a smear toward the smaller fragment sizes.

Shearing DNA

20m

- 3 For good quality DNA: if your DNA extract has fragments that are too long, shear (0.2 mL tube) 500 ng (10 ng/ μ L) of extracted genomic DNA from 200 bp to 500 bp using an ultrasonic cleaner and add 250 μ L of sterilised water at 4°C for 20 minutes.

20m

Testing and adapting the single-tube DNA library preparation by Carøe *et al.* (2018)

5h 38m

4 End-Repair

5m

Start with **14 μ L DNA ($\approx$ 100 ng of DNA)** on a 0.5 μ L LoBind tube. Add **2 μ L of Mastermix** (on ice):

- **0.01 μ L** Taq Polymerase
- **0.01 μ L** T4 DNA Polymerase
- **0.1 μ L** T4 Polynucleotide Kinase
- **1.6 μ L** 10X T4 DNA Ligase Reaction Buffer
- **0.3 μ L** dNTPs (25 mM)



Vortex and centrifuge for 3 seconds.

4.1 Incubate at **20°C** for **30 minutes**.

30m



4.2 Incubate at **65°C** for **30 minutes**.

30m



5 Adapter Ligation

Add to the same tube: **0.7 µL* of Adapter Mix (10 µM, Meyer & Kircher 2010)****. Mix by pipetting.

* In the study of Carøe *et al.* (2018), it is recommended to add 1 µL of adapter mix. However, this was too concentrated for our DNA concentration of <100 ng because we obtained a lot of dimers from the adapters. So, we performed several tests using different volumes of adapter mix (from 1 to 0.3 µL) on butterfly DNA extractions. We found that the best amount of adapter mix is 0.7 µL per <100 ng of total DNA. However, it can be important to adjust this volume on a case-by-case basis according to the DNA quantity.

** For this protocol on butterfly DNA extractions, we decided to use the Meyer & Kircher (2010)'s adapters. In fact, after several tests to compare the Meyer & Kircher's adapters with those used by Carøe *et al.*, we concluded that the Meyer & Kircher's adapters were more efficient since the number of adapter dimers is reduced (results not shown).

5.1 Add **3 µL of Mastermix** (on ice):

5m

- **2.5 µL** Polyethyleneglycol 4000 (50%)
- **0.4 µL** 10X T4 DNA Ligase Reaction Buffer
- **0.1 µL** T4 DNA Ligase



Vortex and centrifuge for 3 seconds.

5.2 Incubate at **20°C** for **1h30 ***

1h 30m

* Instead of 30 min as in Carøe *et al.* (2018).



6 Fill-in

5m

Add to the same tube **10 µL of Mastermix** (on ice):



- **0.3 µL** dNTPs (25 mM)
- **3 µL** Isothermal Amplification Buffer
- **0.5 µL** Bst 2.0 Warmstart Polymerase
- **6.5 µL** Nuclease-Free Water

6.1 Incubate at **65°C** for **20 minutes**.

20m



6.2 Incubate at **80°C** for **20 minutes**.

20m



7 **Sample purification:**

40m



To adapt to the quality of our DNA, we used different purification methods in this protocol.

- In the study by Carøe et al. (2018), they performed the Qiagen protocol of the MinElute PCR Purification Kit. This method is used to purify DNA with short fragments (70kb to 4kb). Follow the manufacturer's instructions. Elute in **30 µL of Nuclease-Free water**.

- For DNA with less fragmentation and larger amounts, we performed Agencourt AMPure XP beads purification according to the manufacturer's instructions (**ratio 2X**). This method is less expensive than the MinElute PCR Purification Kit for a larger number of samples. Elute in **30 µL of Nuclease-Free water**.

- Finally, for DNA with very short fragments and very low concentration, we did not purify and performed the PCR directly.

7.1 For the two first purifications methods, DNA was quantified using the Qubit 4.0 fluorometer.

5m



8 **Indexed PCR**

1m



- For delicate samples without purification, take between 5 and 10 µL of DNA. Fill up with water to a volume of 23 µL.

- For others samples: Take **15 ng** of sample DNA. And fill up with water to a volume of 23 µL.

8.1 Add **1 µL** of Reverse IndexR_01 Primer

1m

8.2 Add **26 µL of Mastermix** (on ice):

- **25 µL** of 2X KAPA HiFi HotStart ReadyMix
- **1 µL** of Forward primer IS4

8.3 Here is the programme we used for the PCRs:

	A	B	C	D
	Step		Temperature	Time
	Initial denaturation		98°C	45 s
	Final extension		72°C	60 s
	Cycling:	Denaturation	98°C	15 s
		Annealing	64°C	30 s
		Extension	72°C	30 s

Table 1. PCR programme with the different steps and time of each step.

Adjust the number of cycles according to the amount of DNA after purification: **12 cycles** for the most concentrated, up to **17 cycles** for the least concentrated.

8.4 Purify the sample using Agencourt AMPure XP beads purification, follow the manufacturer's instructions. Elute in **30 µL of Nuclease-Free water**.

8.5 DNA was quantified using the Qubit 4.0 fluorometer.

Comparison of the single-tube library preparations with different adapters

9 DNA quality was also checked by running a 1% agarose gel.

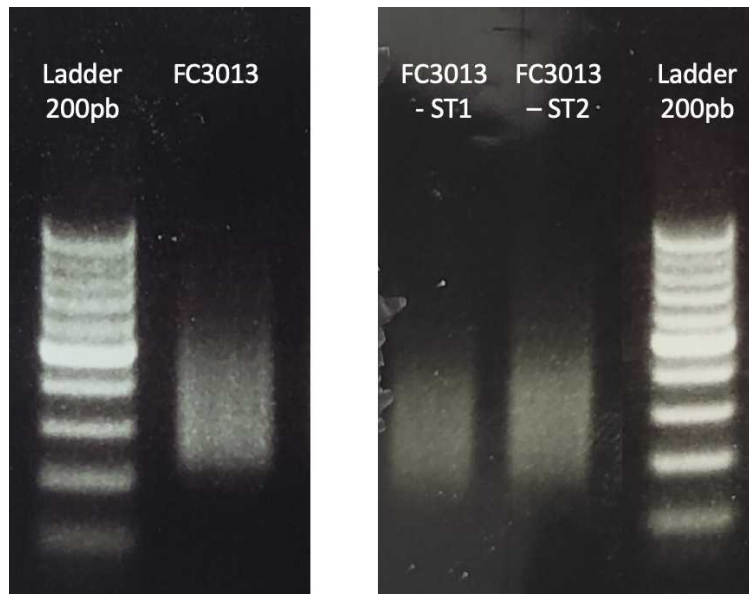


Figure 2. Pictured is an agarose gel showing DNA after PCR amplification of the classical protocol of Tilak *et al.* (FC3013, *left*), and an agarose gel of the single-tube library with Carøe *et al.*'s adapters (FC3013-ST1, *center*) and the single-tube library with Meyer & Kircher's adapters (FC3013-ST2, *right*).

Based on the agarose gel, we found that the three library preparations were similar in size, but with smaller fragment sizes for the single-tube library preparations (**Figure 2**).

Comparison of the raw sequencing results

- 10 We sequenced the three library preparations on the same flow cell on an Illumina NovaSeq 6000 (along with other butterfly libraries). We obtained 18.98 Gb for our classical in-house library (18.07% estimated duplication rate), 20.62 Gb for the single-tube library with Carøe *et al.*'s adapters (19.57% estimated duplication rate), and 20.71 Gb for the single-tube library with Meyer & Kircher's adapters (17.96% estimated duplication rate).

After removing adapters with *fastp* 0.23.2 (Chen *et al.* 2018), we obtained insert size distributions that were similar but not identical. As expected, the single-tube libraries showed a left-shifted distribution indicating a higher proportion of short reads (<100 pb; **Figures 3-5**).

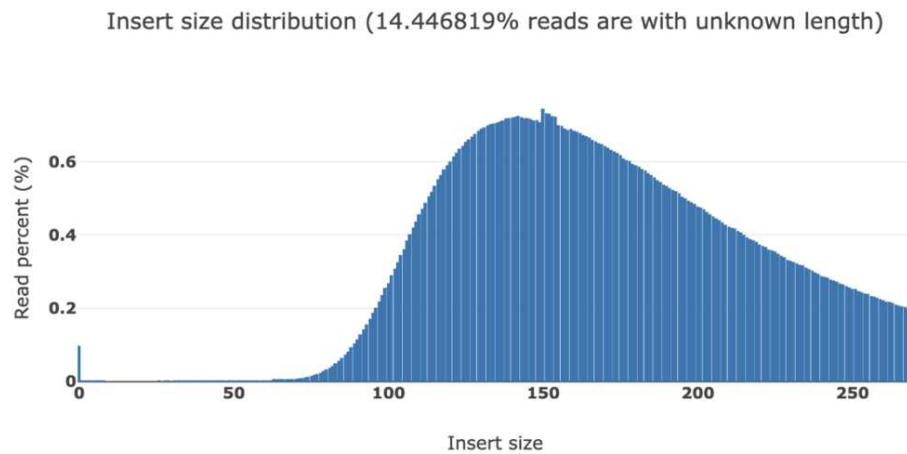


Figure 3. Insert size distribution from *fastp* output for the single-tube library using the protocol of Tilak *et al.* (FC3013).

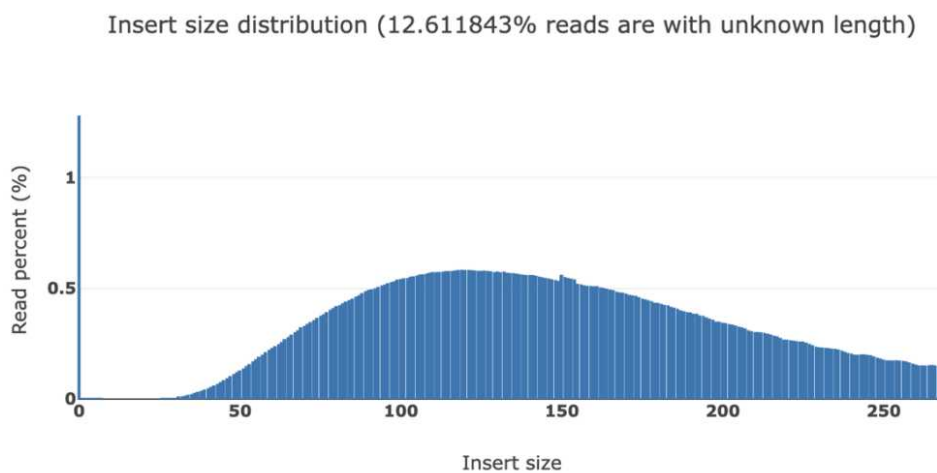


Figure 4. Insert size distribution from *fastp* output for the single-tube library using Carøe *et al.*'s adapters (FC3013-ST1). Note the peak of insert size = 0 that we struggle to explain.

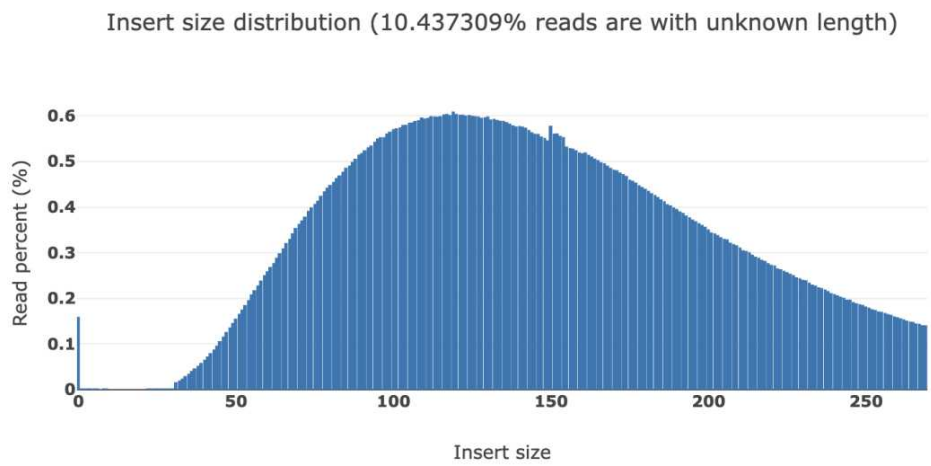


Figure 5. Insert size distribution from *fastp* output for the single-tube library using Meyer & Kircher's adapters (FC3013-ST2).

To validate the single-tube protocol on museum butterflies, we then wondered whether these three libraries could be equally good to assemble the whole genome of the *Papilio machaon* specimen; a species for which a reference genome is already available (Lohse *et al.* 2022: GCA_912999745.1).

Comparison of the genome assemblies

11 To test whether the whole-genome sequencing worked well so that genome assemblies are of good quality to proceed with micro or macroevolutionary analyses, we performed a genome assembly for each library. We ran *MEGAHIT* 1.2.9 (Li *et al.* 2015) with default options, notably the default parameters for the maximum, minimum, and increment of k-mer size in each iteration. We then processed the resulting assemblies through the *gVolante* platform (Nishimura *et al.* 2017) using the *BUSCO* database (Manni *et al.* 2021) to assess genome completeness (*odb_10* database, Lepidoptera dataset, composed of 5,286 orthologous genes).

Based on our results, we found very similar statistics for the three genome assemblies (**Table 2**). The genome assembled with the "classical" in-house library, which we used for our whole-genome sequencing of all swallowtail butterfly species (e.g. Reboud *et al.* 2024), was slightly larger than the two single-tube libraries (+ 8 Mb). However, the genome assembled with the "classical" in-house library was more fragmented (>14,000 scaffolds) and had more missing genes (> +3%) than the genomes assembled from single-tube libraries.

A	B	C	D	E
Library type	Genome size (pb)	Number of	BUSCO score (single complete/duplicated/fr	N50 (pb)

A	B	C	D	E
		scaffolds	agmented/missing)	
Classical in-house	358,549,313	251,917	S:63.2%, D:4.3%, F:17.0%, M:15.5%	2137
Single-tube with Carøe et al.'s adapters	350,796,448	237,913	S:69.1%, D:3.0%, F:15.6%, M:12.3%	2206
Single-tube with Meyer & Kircher's adapters	350,448,437	231,669	S:71.9%, D:3.4%, F:14.6%, M:10.1%	2301

Table 2. Comparison of genome statistics after assembly and *BUSCO* analyses. It shows that the three types of library preparations provided similar quality of genome assembly. The single-tube library with Meyer & Kircher's adapters yielded the "best" assembly in terms of statistics.

The completeness of the genomes assembled from single-tube libraries was higher than with the genome made with the classical library (> +6% of complete orthologous genes). Interestingly, we assembled very similar (if not identical, ~350 Mb) genome size with the single-tube library with Carøe *et al.*'s adapters and the single-tube library with Meyer & Kircher's adapters (**Table 2**).

Conclusion

12 Single-tube library preparation is a promising approach for whole genome sequencing. Originally developed for vertebrates, we have successfully applied and extended it to invertebrates by testing and adapting the protocol of Carøe *et al.* to museum butterflies (and even on beetles) through a range of DNA concentrations going from 0.08 to 30 ng/μL. By comparing a classical in-house protocol to the single-tube protocol with different adapters, we validated the results through analytical comparisons of the quality of genome assemblies. We concluded that the single-tube protocol using Meyer & Kircher's adapters provided the best results for the genome assembly. Although we have tested and adapted this protocol for a single sample for which we could perform multiple tests on the same DNA, we have used this single-tube protocol with very low DNA concentrations (~0.08 ng/μL). For this latter, we successfully constructed and sequenced the DNA library and assembled a 'correct' genome with >54% of the orthologous genes from *BUSCO* recovered, despite the large genome size of the butterfly species (1 Gb). This cost- and time-effective protocol paves the way for the sequencing of insect hDNA, in particular coming from museum collections.

Protocol references

- Carøe C., Gopalakrishnan S., Vinner L., Mak S.S., Sinding M.H.S., Samaniego J.A., Wales N., Sicheritz-Pontén T. & Gilbert M.T.P. (2018) Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution*, 9, 410–419.
- Chen S., Zhou Y., Chen Y. & Gu J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34, i884–i890.
- Li D., Liu C.M., Luo R., Sadakane K. & Lam T.W. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, 31, 1674–1676.
- Lohse K., Hayward A., Laetsch D.R., Vila R., Yumnam T. & Darwin Tree of Life Consortium. (2022) The genome sequence of the common yellow swallowtail, *Papilio machaon* (Linnaeus, 1758). *Wellcome Open Research*, 7, 261.
- Manni M., Berkeley M.R., Seppey M. & Zdobnov E.M. (2021) BUSCO: assessing genomic data quality and beyond. *Current Protocols*, 1, e323.
- Meyer M. & Kircher M. (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 6, pdb.prot5448.
- Neiman M., Sundling S., Grönberg H., Hall P., Czene K., Lindberg J. & Klevebring D. (2012) Library preparation and multiplex capture for massive parallel sequencing applications made efficient and easy. *PLoS One*, 7, e48616.
- Nishimura O., Hara Y. & Kuraku S. (2017) gVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics*, 33, 3635–3637.
- Reboud E.L., Nabholz B., Chevalier E., Lafon B.J., Tilak M.-K., Mielke C.G., Cotton A.M. & Condamine F.L. (2024) Clarifying the phylogeny and systematics of the recalcitrant tribe Leptocircini (Lepidoptera: Papilionidae) with whole-genome data. *Systematic Entomology*. <https://doi.org/10.1111/syen.12661>
- Tilak M.-K., Justy F., Debiais-Thibaud M., Botero-Castro F., Delsuc F. & Douzery E.J.P. (2015) A cost-effective straightforward protocol for shotgun Illumina libraries designed to assemble complete mitogenomes from non-model species. *Conservation Genetics Resources*, 7, 37–40.
- Zheng Z., Advani A., Melefors O., Glavas S., Nordström H., Ye W., Engstrand L. & Andersson A.F. (2011) Titration-free 454 sequencing using Y adapters. *Nature Protocols*, 6, 1367–1376.

Acknowledgements

We thank Rodolphe Rougerie and Jérôme Barbut for access to the collections of the Muséum National d'Histoire Naturelle in Paris (France) and Gerardo Lamas for access to the collections of the Museo de Historia Natural in Lima (Peru).