Duchen P., Höhna S. & Catalán A., 2020. Genomics of European fireflies: first results and next steps. *Bulletin de la Société Vaudoise des Sciences Naturelles* 99 : 67-72.

Materials and Methods

Collection of specimens

To start setting up a latitudinal cline we selected the following locations : Arco, Cologno al Serio and Turin in Italy (44.5-45.07º N), Lausanne (46.52º N), and Munich (48.14º N). Species incidence was revised in iNaturalist (https://www.inaturalist.org/) and the Centre Suisse de Cartographie de la Faune (www.cscf.ch). The Italian populations, which lie south of the Alps, provide a great opportunity to study adaptation to warmer climates and the migration capabilities across mountain ranges (figure 1). Additionally, populations from Switzerland constitute an ideal stepping-stone along this cline. Field collections began in late June 2020 and lasted until late July 2020.

Genomic DNA (gDNA) and RNA extraction

Collected specimens for DNA extraction were preserved in ethanol 96 % to ensure DNA stability. For RNA extraction, two to three specimens per firefly species were dissected and preserved in DNA/RNA Shield (Zymo Research). To optimize DNA extraction, four extraction methods were tested : two for high molecular DNA (MagAttract HMW DNA and Genomic-tip 500/G, Qiagen), and two for regular DNA extraction (NEB Monarch No. T3010 and Qiagen DNAeasy No. 69504). The extraction of high molecular DNA is useful for sequencing with long-range technologies and achieve a better genome continuity with higher N50 values. gDNA quality and integrity was checked by gel electrophoresis and Qubit.

Planned whole genome sequencing

Based on the amount of gDNA generated in the previous step, we have chosen to generate long read fragments by sequencing with the promethION platform from Oxford nanopore. This technology will allow us to reach a N50 of 33 kb, which will ensure high continuity during the genome assembly. To correct for base call errors during the promethION run, we will also generate high coverage Illumina reads (60x sequencing depth). The generation of Illumina reads will allow us to polish and correct the generated promethION reads. Furthermore, we will generate a chromosome level assembly using Hi-C Dovetail technologies. Generating a chromosome level assembly will allow us to assign the assembled contigs and scaffolds to chromosomes.

Planned genome *de-novo* assembly and annotation

For the annotation step, we will generate whole transcript sequences using a PacBio Iso-seq approach. Gene prediction and annotation will be done with MAKER (Cantarel *et al.* 2008). Additionally, annotation information from *P. pyralis* (Fallon *et al.* 2018) and NCBI will be leveraged, and BLAT will be used to infer annotation based on orthology (Bhagwat *et al.* 2012). Genome assembly and annotation completeness will be assessed with BUSCO (Simão *et al.* 2015). The availability of whole genome data of *L. noctiluca*, *L. splendidula*, *Luciola italica*, *P. pyralis*, and other firefly genomes such as *Pyrocoelia pectoralis* (Fu 2017), *Aquatica*

lateralis, and the closely related click beetle *Ignelater luminosus* (Fallon et al. 2018) will open the possibility for comparative genomic studies.

Planned whole-genome re-sequencing

To gather polymorphism data from the rest of the localities we need to re-sequence the collected individuals from those populations. For each of the collected populations, Illumina paired-end 150bp-long reads will be generated with a 20x sequencing depth. In addition to populations of *L. noctiluca*, *L. splendidula*, and *Luciola italica*, we will also generate short read sequences of *Lampyris iberica*, which will serve as an outgroup to *L. noctiluca*. Likewise, short read sequences will be generated for *Luciola lusitanica* and *Lamprohiza germari*, which will serve as outgroups for *L. italica* and *L. spendidula*.

Whole genomic DNA will be extracted as described above (section *Genomic DNA (gDNA) extraction*). The read mapper BWA (Li & Durbin 2009) will be used to map the generated reads to the corresponding reference genome. The mapped files in BAM format will be further curated by removing PCR duplicates. Likewise, low quality mapped reads will be curated using SAMtools (Li *et al.* 2009). Nucleotide polymorphism diversity estimates will be estimated with VCFtools (Danecek *et al.* 2011), and variant calling will be done with ATLAS (Link *et al.* 2017). To call ancestral and derived states, outgroup sequences will be mapped to their corresponding reference genome.

References

Bhagwat M., Young L., & Robison R.R., 2012. Using BLAT to find sequence similarity in closely related genomes. *Current protocols in Bioinformatics* 37: 10-8.

Cantarel B.L., Korf I., Robb S.M., Parra G., Ross E., Moore B., Holt C., Alvarado A. S., & Yandell M., 2008. Maker: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome research* 18: 188-196.

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., *et al.*, 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156-2158.

Fallon T.R., Lower S.E., Chang C.-H., Bessho-Uehara M, Martin G.J., Bewick A.J., Behringer M., Debat H.J., Wong I., Day J.C., *et al.*, 2018. Firefly genomes illuminate parallel origins of bioluminescence in beetles. *elife* 7: e36495.

Fu X., Li J., Tian Y., Quan W., Zhang S., Liu Q., Liang F, Zhu X., Zhang L., Wang D., *et al.*, 2017. Long-read sequence assembly of the firefly *Pyrocoelia pectoralis* genome. *GigaScience* 6: gix112.

Li H. & Durbin R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25: 1754-1760.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., & Durbin R., 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078-2079.

Link V., Kousathanas A., Veeramah K., Sell C., Scheu A., & Wegmann D., 2017. ATLAS: analysis tools for low-depth and ancient samples. *bioRxiv*: 105346.

Simão F.A., Waterhouse R.M, Ioannidis P., Kriventseva E.V., & Zdobnov E.M., 2015. Busco: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210-3212.