Erythrinus erythrinus samples from 2 different karyomorphs were collected in 3 different locations. 6 individuals of karyomorph A, code A1 collected in 16.1713/-55.9573. 5 individuals of karyomorph A code A2 collected in -23.3774/-53.7805. 7 individuals of karyomorph D, code D collected in -5.7770/-35.2092.

**Methods**

The DArTseq sequencing procedure were performed using liver tissues, that were shipped to Diversity Arrays Technology Pty Ltd (Australia). This method involves a complexity-reduction in which the DNA is treated with PstI and SbfI enzymes before sequencing, which might enrich for sequences in lightly methylated regions (Killian et al. 2018). Sequencing was carried out on an Illumina HiSeq 2500 platform. The raw data processing and all subsequent steps to obtain the final datasets were executed using pyRAD v3.0.66  (Eaton, 2014). The minimal depth of coverage to form a cluster was set to 6, the clustering threshold implemented was 0.88. On the last alignment step, the minimal sample coverage for a locus was set to 18, therefore only data present in all individuals were maintained. The sequencing adapters were trimmed and sequences with more than five low-quality bases were removed (Q < 20). Subsequently, reads were aligned, and consensus sequences were obtained for each sample, which were used to define the reference base at each position in each alignment. We obtained loci that were present in all sampled individuals and generated two different datasets also in pyRAD v3.0.66 (Eaton, 2014) for further analysis: (1) a phased sequence file for each loci that was implemented only in the genetic diversity and differentiation analyzes based in the “.alleles format” and (2) a matrix of polymorphic single nucleotide polymorphisms (SNPs), with a single SNP per locus to prevent the presence of closely linked SNPs based in the “.usnps.geno” output. The SNP matrix was coded as follows: 0 for homozygotes for the reference base, 1 for heterozygotes, and 2 for the alternate base homozygotes. Indels were not considered, and this second input was used in the remaining analyzes.

Samples number and codes are indicated below, the order of the samples in the Dataset2 Archive is the same as here presented:

|  |  |
| --- | --- |
| 2077489 | A2 |
| 2077490 | D |
| 2077497 | A2 |
| 2077498 | D |
| 2077505 | A2 |
| 2077506 | D |
| 2077513 | A2 |
| 2077514 | D |
| 2077521 | A2 |
| 2077522 | D |
| 2077529 | A1 |
| 2077537 | A1 |
| 2077545 | A1 |
| 2077546 | A2 |
| 2077553 | A1 |
| 2077554 | D |
| 2077561 | A1 |
| 2077569 | A1 |

**Funding**

Fundação de Amparo à Pesquisa do Estado de São Paulo, Award: 2019/25009-7

Fundação de Amparo à Pesquisa do Estado de São Paulo, Award: 2020/11772-8

Conselho Nacional de Desenvolvimento Científico e Tecnológico, Award: 302449/2018-3