Assembly and annotation of the sable *(Martes zibellina)* and pine marten *(Martes martes)* genomes

Authors: Andrey A. Tomarovsky [1], Azamat A. Totikov [1], Violetta R. Beklemisheva [2], Polina L. Perelman [2], Natalia A. Serdyokova [2], Tatiana Bulyonkova [3], Ksenia A. Koniaeva [4], Alexei V. Abramov [5], Alexander S. Graphodatsky [2], Klaus-Peter Koepfli [6], Roger A. Powell [7], Sergei F. Kliver [8].

[1] Computer Technologies Laboratory, ITMO University, 49 Kronverkskiy Pr., 197101 Saint Petersburg, Russia; andrey.tomarovsky@gmail.com; a.totickov1@gmail.com
[2] Institute of Molecular and Cellular Biology SB RAS, 8/2 Acad. Lavrentiev Ave, 630090 Novosibirsk, Russia, bekl@mcb.nsc.ru; perelmanp@mcb.nsc.ru; serd@mcb.nsc.ru
[3] A.P. Ershov Institute of Informatics Systems SB RAS, 6 Acad. Lavrentiev Ave, 630090 Novosibirsk, Russia, ressaure@gmail.com
[4] Barnaul Zoo "Lesnaia skazka", 12 Enthuziastov str, 630090 Barnaul, Russia, men1965@yandex.ru
[5] Laboratory for Theriology, Zoological Institute RAS, 1 Universitetskaya Emb, 199034 Saint Petersburg, Russia, a.abramov@mail.ru
[6] Smithsonian-Mason School of Conservation, 1500 Remount Road, 22630 Front Royal, VA, USA, klauspeter.koepfli527@gmail.com
[7] Department of Applied Ecology, North Carolina State University, Campus Box 7617, 27695-7617 Raleigh, NC, USA, newf@ncsu.edu
[8] Institute of Molecular and Cellular Biology SB RAS, 8/2 Acad. Lavrentiev Ave. 8/2, Novosibirsk 630090, Russia, mahajrod@gmail.com

Introduction

The sable (*Martes zibellina*) and the pine marten (*Martes martes*) are closely-related but distinct species. In their sympatric zone, they breed and produce hybrids called "kidases". Kidases have intermediate phenotypes, with traits distinguishing them from both the sable and pine marten. However, very little is known about the genetic background and level of admixture of these species in the hybridization area. Only low resolution methods such as STRtyping or mtDNA-based approaches were previously used to study these problem.

The usual bottleneck in conservation genomics is the generation of a reference genome of suitable quality. The current standard even in conservation biology is chromosome level assembly. Yet no such assembly has been published for any of the species of the Guloninae lineage; only two highly fragmented draft assemblies based on obsolete jumping library approach are available for the wolverine (Gulo gulo) and the sable. In this study, we present the first de novo chromosome length assemblies for the sable and the pine marten, and a primary comparison of the two species.

Methods

The reference genome assemblies for the sable and the pine marten were generated using linked read (10X) and HiC-sequencing. Additional individuals from the sympatric zone were resequenced with a ~30x coverage on the Illumina platform. Initial annotation included identification of repetitive elements, prediction of both protein-coding and ncRNA genes, and localization of the pseudoautosomal region (PAR) on the X chromosome. A custom script implementing coverage-based approach was written for PAR detection. The genomes of the two species were compared using whole genome alignment.

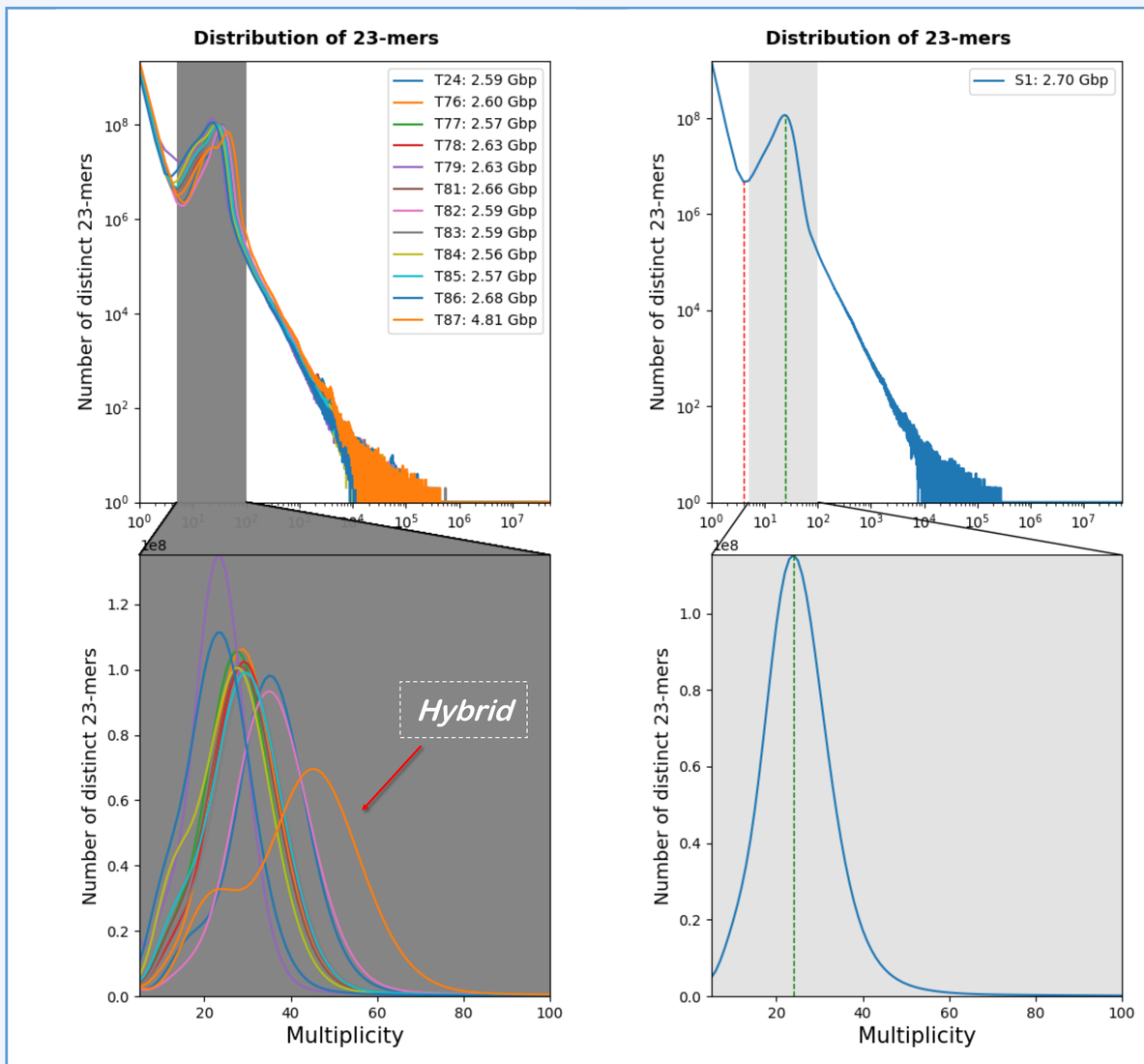

Figure 2. 23-mers distributions for potential kidases and sable sample

Conclusions

Chromosomal length assemblies of sable and pine marten genomes closed a white spot in the genomics of Guloninae lineage. Assemblies of such type greatly simplify analysis and visualization of genetic diversity and help in solving other problems in conservation genomics. Our efforts now are directed at elucidating the genetic background of hybridization between different marten species, and we have already made the first step in this direction by analysing kidas-like individuals from the zone of sympatry.

Results and Discussions

We generated high-quality chromosome-length reference genome assemblies for *M. zibellina* and *M. martes*. Both assemblies have approximately the same length of 2.4 Gbp and a high N50: 143 Mbp and 144 Mbp for the sable and the pine marten, respectively (Table 1). Moreover, the number of chromosomal scaffolds in the assemblies corresponds to the haploid number of chromosomes for both species (1n=19).

Latin name	Common name	Assembly type*	2n	Length, Gbp	Ns, Mbp	N50	L50
Martes Martes	Pine marten	HiC	38	2.4	25.2	144638580	7
Martes zibellina	Sable	HiC	38	2.4	25.1	143641668	7

Table 1. *M. martes* and *M. zibellina* genome assemblies used in this study.
* Assembly type: HiC - chromosome-level assembly.

BUSCO-based assessment also indicated high completeness of the obtained sequences, with 95,56% complete BUSCOs in each assembly (Table 2). Venn diagrams were drawn based on the results (Figure 1).

BUSCO status	*M. martes*	*M. zibellina*
Complete	3922	3922
Complete and single-copy	3895	3901
Complete and duplicated	27	21
Fragmented	95	101
Missing	87	81
Total	4104	4104

Table 2. Results of assessing the quality of genomic assemblies by conservative orthologs (BUSCOs)

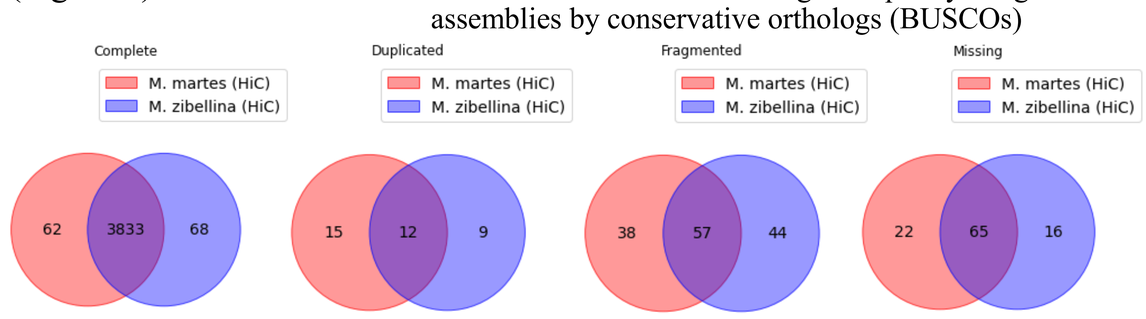

Figure 1. Venn diagrams of overlapping of different BUSCO categories for chromosome assemblies of sable and pine marten

Potential kidas-like individuals were resequenced with a 30x coverage. Preliminary analysis of 23-mer distributions extracted from the corresponding reads indicated the presence of at least one candidate F1 hybrid (Figure 2).

PAR length was the same for all individuals, including the reference ones. The coordinates of the pseudoautosomal region were assessed with a precision of 5 kbp. However, the orientation of the X chromosome is different in the sable and the pine marten assemblies. In the sable assembly, PAR is located in the beginning (0.145 Mbp – 6.675 Mbp) of the X chromosome C-scaffold, while in the pine marten assembly, it is located on the opposite end (118.000 Mbp – 124.585 Mbp). Whole genome alignment of genome assemblies indicated opposite orientation of several other C-scaffolds (Figure 3).

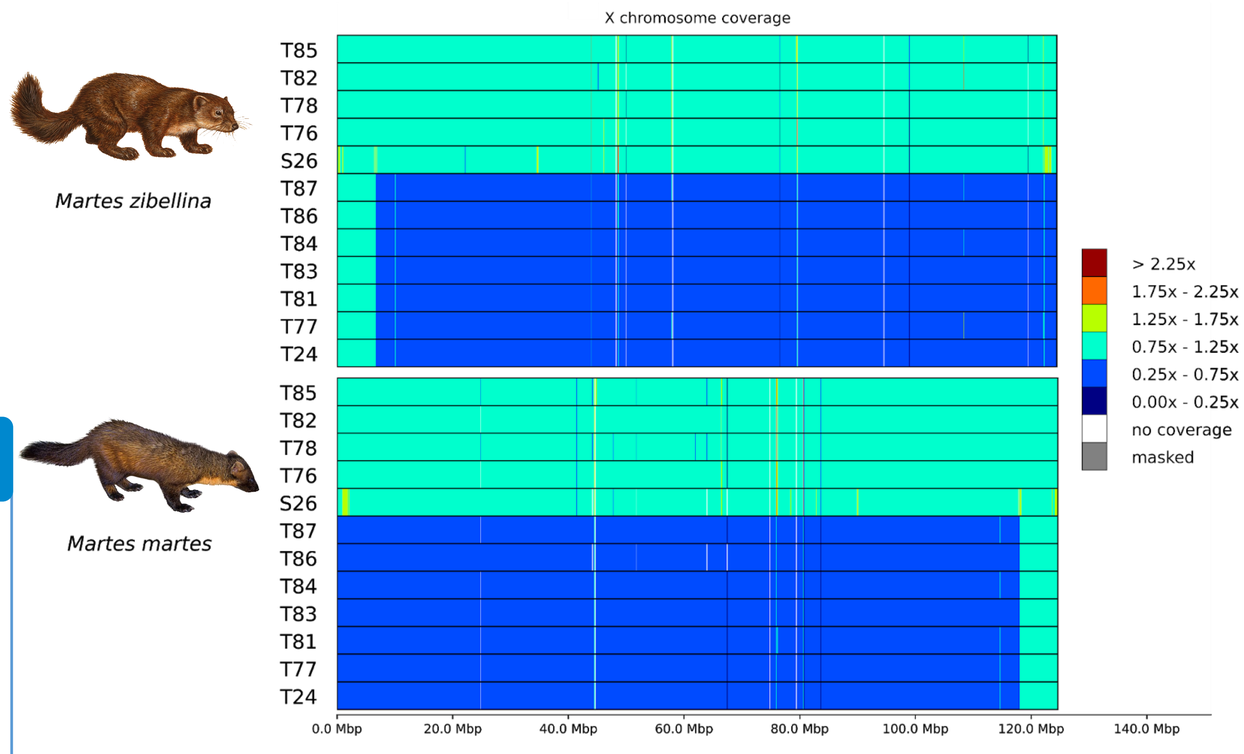

Figure 1. Coverage of the X chromosome of *M. zibellina* and *M. martes* assemblies

Acknowledgments: We are grateful to Sergei Pisarev (Barnaul Zoo) and Alexander Migura for their help with sample gathering. The reported study was funded by JetBrains Research and the Russian Foundation for Basic Research (RFBR), project number 20-04-00808.