

GENOME SEQUENCING

Chromosome-scale genome sequences of two Charadriiformes species: *Scolopax rusticola* and *Lymnocyrtes minimus*

Emilie Delpuech¹, Maeva Leitwein¹, Arnaud Couloux², Karine Labadie³, Corinne Cruaud³, Jean-Marc Aury², Carole Bodin⁴, Pascal Lapébie⁴

¹ LDgenX, 34690 Fabrègues, France, ² Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France, ³ Genoscope, Institut François Jacob, CEA, Université Paris-Saclay, 91057 Evry, France, ⁴ Fédération Nationale des Chasseurs, 92136 Issy les Moulineaux Cedex, France

<https://doi.org/10.56179/001c.143981>

Biodiversity Genomes

We present high-quality genome assemblies of two snipes: the Eurasian Woodcock (*Scolopax rusticola*) and the Jack Snipe (*Lymnocyrtes minimus*). Illumina and Nanopore sequencing were performed on fresh blood samples of wild specimens. The reads were assembled using a *de novo* method followed by the integration of Hi-C data to reconstruct chromosomal structures. Raw and assembled data are publicly available in GenBank.

Introduction

The Eurasian Woodcock (*Scolopax rusticola*) and the Jack Snipe (*Lymnocyrtes minimus*) are two Charadriiformes belonging to the *Scolopacidae*.

The Eurasian Woodcock is one of the eight species of the *Scolopax* genus (Winkler, Billerman, and Lovette 2020) distributed along the Palearctic. The Eurasian Woodcock is unique as it is the only member of the wading bird family that is not dependent on wetlands and instead lives in forested areas. This is a discreet bird mainly active at night. The species is migratory and divided into two populations, the European and the Asian. The species is listed by the IUCN (2024) as at least concern and hunted in most countries.

The Jack Snipe (*Lymnocyrtes minimus*) is the only member of the genus *Lymnocyrtes* (Winkler, Billerman, and Lovette 2020). Jack snipe is a migrating species breeding in North Europe (Scandinavia et Siberia) and wintering in Middle East and Europe (Lepley et al. 2005; Battisti and Soldato 2018). The Jack snipe is a discreet and nocturnal bird occupying wetland habitat such as petland and swamp. The species is also listed by the IUCN (2024) as at least concern and mainly hunted in Europe and Ireland.

Methods

Blood from a single wild *Scolopax rusticola* individual was collected during winter 2021-2022 in northeastern France. The bird belong to the Europe/South & West Europe & North Africa population (AEWA).

Blood from a single wild *Lymnocyrtus minimus* was collected during winter 2021-2022 in northern France. The bird belong to the Northern Europe/S & W Europe & West Africa population (AEWA).

DNA preparations, sequencing and assembly analysis were performed by Genoscope (France).

Library preparation and sequencing

DNA extractions were performed on blood using either the NucleoBond HMW DNA Extraction Kit (Macherey-Nagel, Düren, Germany) or Nanobind CBB kit (Pacific Biosciences, CA, USA), following the manufacturer's instructions. DNA was quantified by a dsDNA-specific fluorimetric quantitation method using Qubit dsDNA HS Assays (ThermoFisher Scientific, Waltham, MA). High molecular weight (HMW) gDNA quality was checked on a Femto Pulse system (Agilent, CA, USA) and the mean length of the DNA molecules was estimated to be over 90 Kb. HMW gDNA was first size-selected using Short Read Eliminator kit (Pacific Biosciences, CA, USA).

Long Read libraries were prepared using the Oxford Nanopore SQK-LSK114 kit and loaded on R10.4.1 PromethION flow cells, producing over 70 Gb of data per genome. The nanopore long reads were not cleaned and raw reads were used for genome assembly.

PCR free libraries were prepared following the Kapa Hyper Prep Kit procedure provided by Roche (Roche, Basel, Switzerland). They were sequenced on a paired-end mode with an Illumina NovaSeq 6000 instrument (Illumina, San Diego, CA, USA) using a 151 base-length read chemistry. Short Illumina reads were bioinformatically post-processed sensu Alberti et al. (2017) to filter out low-quality data.

Omni-C libraries were prepared using the Dovetail Omni-C Kit (Dovetail Genomics, Scotts Valley, CA, USA), following the OmniC Proximity Ligation Assay Non Mammalian Samples Protocol for nucleated blood (version1.0). Libraries were sequenced on the NovaSeq6000, a sequencing platform from Illumina (San Diego, CA, USA) with 2x 150 read length, and more than 170 million Omni-C reads were generated for each genome.

RNA extractions were performed using the RNeasy Plus Universal kit (Qiagen, Hilde, Germany). RNA libraries were prepared following the Illumina Stranded mRNA Prep, Ligation kit (Illumina, San Diego, CA, USA) and were sequenced with the Illumina NovaSeq6000 using a paired-end 151 base-length read chemistry. Short Illumina reads were bioinformatically post-processed sensu Alberti et al. (2017) to filter out low-quality data.

Long reads-based genome assembly

The *Lymnocyptes minimus* and *Scolopax rusticola* genome sizes were estimated using Genomescope (Ranallo-Benavidez et al. 2020), Illumina reads, and a k-mer value of 31.

To generate long-reads based genome assemblies we generated three samples of reads: (i) all read, (ii) 30X coverage of the longest reads, and (iii) 30X coverage of the filtlong (Wick 2022) highest-score reads that were used as input data for four different assemblers, Smartdenovo (Liu et al. 2021), Redbean (Ruan and Li 2020), Flye (Kolmogorov et al. 2019) and Necat. Smartdenovo was launched with -k 17 and -c 1 to generate a consensus sequence. Redbean was launched with '-xont -X5000 -g1200m' and Flye with '-g 1200m'. The resulting assemblies were evaluated based on cumulative size and contiguity, with the Smartdenovo and all-reads combination producing the best assembly for *Lymnocyptes minimus*, and the Smartdenovo and filtlong reads combination for *Scolopax rusticola*. These two assemblies were polished using Medaka ("Medaka" 2021) with Nanopore reads and two times with Hapo-G (Aury and Istace 2021) and Illumina PCR-free reads (Table S1).

Then, both assemblies were scaffolded with Hi-C data using YaHS (Zhou, McCarthy, and Durbin 2023, 2023). The assemblies were checked for contamination and manually curated using Juicebox (Durand et al. 2016) (Figures S1 and S2).

Results and Data Availability

All raw read data and assembled genomes files are available in the GenBank database (Table 1).

Table 1. Accession number and Bioproject of both *L. minimus* and *S. rusticola*.

Species	Accession number	Sequencing data	Assembled genomes
<i>Scolopax rusticola</i>	PRJEB78459	PRJEB77913	PRJEB78438
<i>Lymnocyptes minimus</i>	PRJEB78460	PRJEB77912	PRJEB78437

The estimated genome size was 1,257Mb and 1,098Mb for *L. minimus* and *S. rusticola* (Figure S3) with a low heterozygosity rate (0.546% and 0.788%). Busco scores indicates a high-quality complete assembly with an average of 98.6% across eukaryotic databases (Table S1).

Contribution to the work

This work was part of the Limicomics project, funded by the Fédération Nationale des Chasseurs (FNC).

CB, PL conceptualized, supervised and administrated the study, organized the collect and provided the blood samples. They reviewed and edited the manuscript.

AC, KL, JMA and CC performed the DNA preparation, sequencing and assembly analysis, reviewed and edited the manuscript. (Genoscope is a platform of the Commissariat à l'Energie Atomique (CEA) and is part of the France Génomique Structure (ANR-10-INBS-09-08)).

ED and ML wrote the original draft, reviewed and edited the manuscript.

Acknowledgement

We are grateful to Christophe Urbaniak, and Benoit Balangé from the Société de Chasse Militaire de Mourmelon. We also thank the Club International des Chasseurs de Bécassines (CICB), Jeanne Abbou and Mickaël Mourrier from the Hutte des 400 coups.

Submitted: July 19, 2025 EDT. Accepted: July 19, 2025 EDT. Published: September 04, 2025 EDT.

REFERENCES

- Alberti, Adriana, Julie Poulain, Stefan Engelen, Karine Labadie, Sarah Romac, Isabel Ferrera, Guillaume Albin, et al. 2017. "Viral to Metazoan Marine Plankton Nucleotide Sequences from the Tara Oceans Expedition." *Scientific Data* 4 (1): 170093. <https://doi.org/10.1038/sdata.2017.93>.
- Aury, Jean-Marc, and Benjamin Istace. 2021. "Hapo-G, Haplotype-Aware Polishing of Genome Assemblies with Accurate Reads." *NAR Genomics and Bioinformatics* 3 (2): lqab034. <https://doi.org/10.1093/nargab/lqab034>.
- Battisti, Andrea, and Giovanni Soldato. 2018. "Regular Wintering of Jack Snipe (*Lymnocyrtus Minimus*) in NW Italian Heathland Habitat." *Rivista Italiana Di Ornitologia* 88 (1): 3–8. <https://doi.org/10.4081/rio.2018.334>.
- Durand, Neva C., James T. Robinson, Muhammad S. Shamim, Ido Machol, Jill P. Mesirov, Eric S. Lander, and Erez Lieberman Aiden. 2016. "Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom." *Cell Systems* 3 (1): 99–101. <https://doi.org/10.1016/j.cels.2015.07.012>.
- IUCN. 2024. "The IUCN Red List of Threatened Species. Version 2024-2." <https://www.iucnredlist.org>.
- Kolmogorov, Mikhail, Jeffrey Yuan, Yu Lin, and Pavel A. Pevzner. 2019. "Assembly of Long, Error-Prone Reads Using Repeat Graphs." *Nature Biotechnology* 37 (5): 540–46. <https://doi.org/10.1038/s41587-019-0072-8>.
- Lepley, Michel, Pierre Defos Du Rau, Mickaël Veillé, Olivier Pineau, and Jean Yves Mondain Monval. 2005. "Capturing Jack Snipe *Lymnocyrtus Minimus* with Mobile Horizontally Held Nets." *Ringling & Migration* 22 (3): 167–70. <https://doi.org/10.1080/03078698.2005.9674327>.
- "Medaka." 2021. Oxford Nanopore Technologies. <https://github.com/nanoporetech/medaka>.
- Ranallo-Benavidez, T., Rhyker T., Kamil S. Jaron, and Michael C. Schatz. 2020. "GenomeScope 2.0 and Smudgeplot for Reference-Free Profiling of Polyploid Genomes." *Nature Communications* 11 (1): 1432. <https://doi.org/10.1038/s41467-020-14998-3>.
- Ruan, Jue, and Heng Li. 2020. "Fast and Accurate Long-Read Assembly with Wtdbg2." *Nature Methods* 17 (2): 155–58. <https://doi.org/10.1038/s41592-019-0669-3>.
- Wick, R. 2022. "Filtlong." <https://github.com/rrwick/Filtlong>.
- Winkler, David W., Shawn M. Billerman, and Irby J. Lovette. 2020. "Sandpipers and Allies (Scolopacidae), Version 1.0." *Birds of the World*. https://doi.org/10.2173/bow.scolop2.01species_shared.bow.project_name.
- Zhou, Chenxi, Shane A. McCarthy, and Richard Durbin. 2023. "YaHS: Yet Another Hi-C Scaffolding Tool." *Bioinformatics* 39 (1): btac808. <https://doi.org/10.1093/bioinformatics/btac808>.

Table S1. General statistics of the *Scolopax rusticola* and *Lymnocyrtus minimus* genome assemblies

	<i>Scolopax rusticola</i> (Flye-Filtlong)	<i>Lymnocyrtus minimus</i> (Fly-All)
Cumulative size	1,146 Mb	1,218Mb
N50 (L50)	23 Mb (16)	22 Mb (18)
N90 (L90)	4 Mb (57)	3 Mb (69)
Number of contigs	547	1 290
Busco score	Euk : C:98.8%[S:98.4%,D:0.4%],F:1.2%,M:0.0%,n:255 Aves : C:97.3%[S:96.6%,D:0.7%],F:0.6%,M:2.1%,n:8338	Euk : C:98.4%[S:97.6%,D:0.8%],F:1.2%,M:0.4%,n:255 Aves : C:97.5%[S:97.0%,D:0.5%],F:0.5%,M:2.0%,n:8338

Figure S1. Hi-C map for the *Lymnocyrtus minimus* genome assembly.

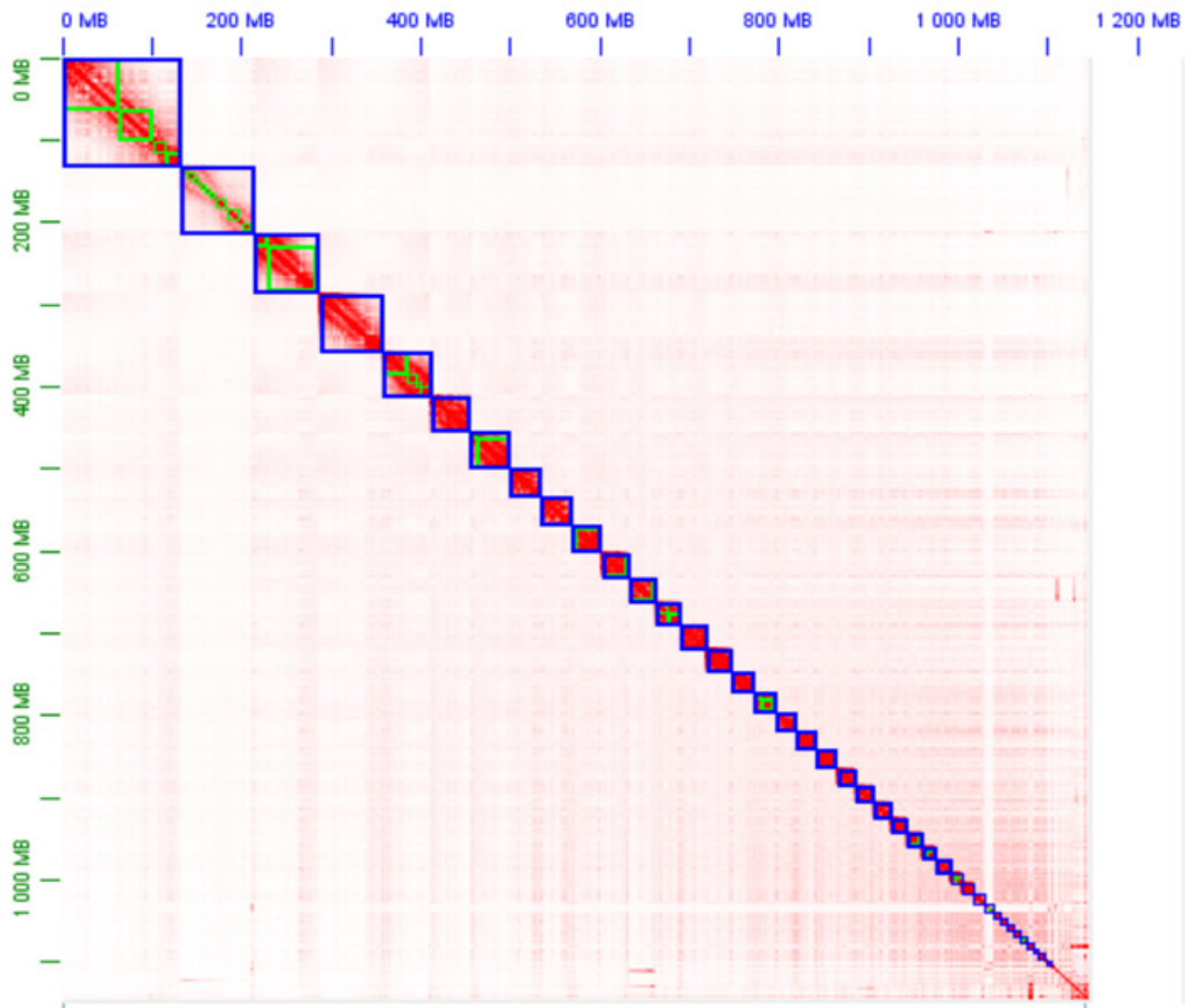
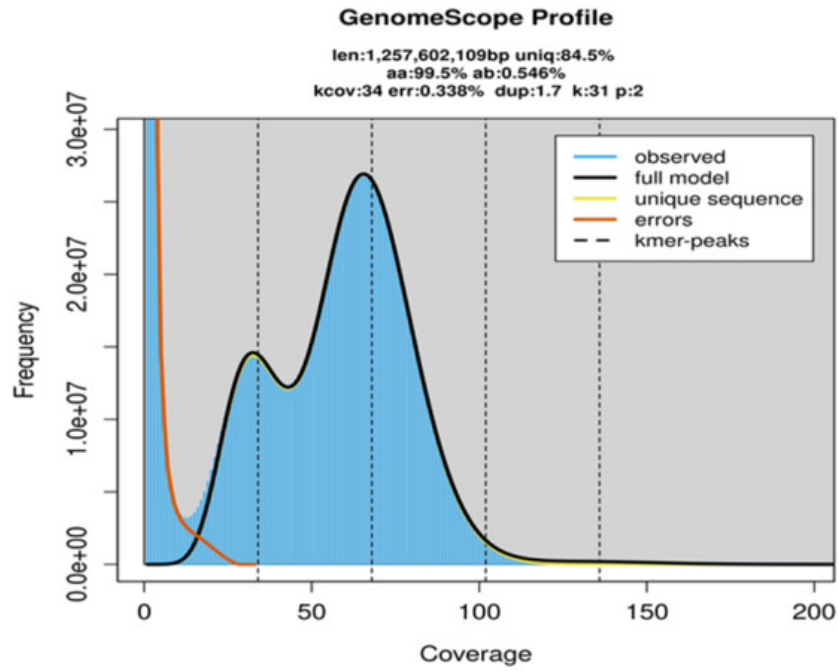


Figure S2. Hi-C map for the *Scolopax rusticola* genome assembly.

A.



B.

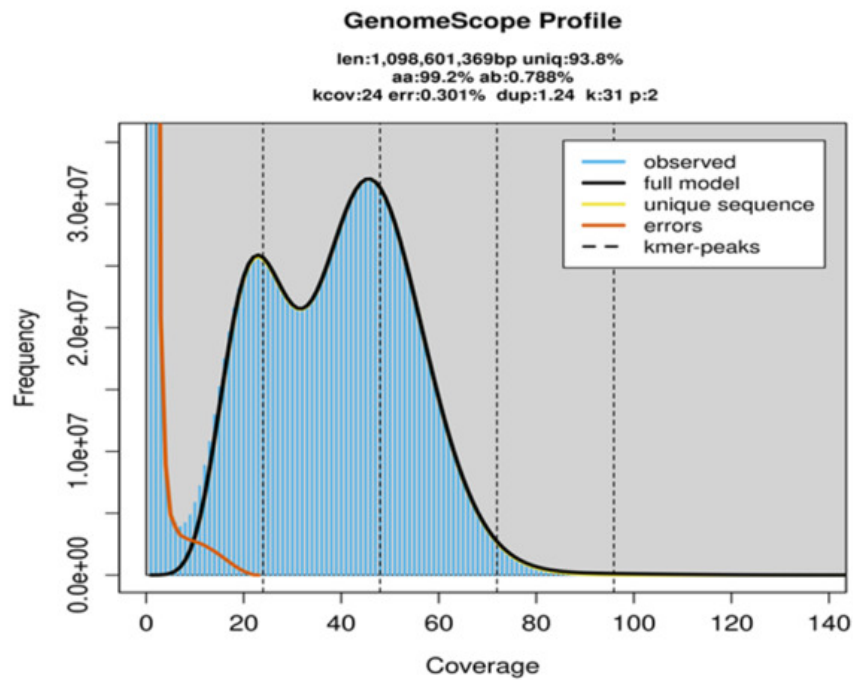


Figure S3. Genomescope outputs for the **A.** *Lymnocyrtus minimus* and **B.** *Scolopax rusticola* genomes.