

Data References Table

from "Interpreting the pervasive observation of U-shaped Site
Frequency Spectra"

Fabian Freund, Elise Kerdoncuff, Sebastian Matuszewski, Marguerite Lapierre
Marcel Hildebrandt, Jeffrey D. Jensen, Luca Ferretti, Amaury Lambert,
Timothy B. Sackton and Guillaume Achaz

February 2, 2023

Corresponding author: guillaume.achaz@mnhn.fr

Table 1: Data set information

Species	Outgroup	<i>n</i>	Outgroup size	Polarized SNPs	≠ outgroup	Diallelic# outgroup	Source	Outgroup source (or same as sample)
<i>Acinetobacter baumannii</i>	<i>A. nosocomialis</i>	79	1	78175	6006		[RdSB ⁺ 18] [*]	NCBI RefSeq● - Nov 2016
<i>Aptenodytes patagonicus</i>	<i>A. forsteri</i>	20	2	1278	12	32	[RGB ⁺ 14] ⁺	[FBRG14]
<i>Arabidopsis thaliana</i>	<i>A. lyrata</i>	345	1	10322757	1023148	398365	[ABAB ⁺ 16]	Public directory○
<i>Armadillidium vulgare</i>	<i>A. nasatum</i>	20	4	23323	745		[RGB ⁺ 14] ⁺	same
<i>Artemia franciscana</i>	<i>A. sinica</i>	20	4	5548	247		[RGB ⁺ 14] ⁺	same
<i>Athene cunicularia</i>	<i>Strix occidentalis</i>	40	2	11268203	383702	68196	[MKB ⁺ 18] [*]	[HHW ⁺ 17]
<i>Bacillus subtilis</i>	<i>B. atrophaeus</i>	38	1	105523	29934		[RdSB ⁺ 18] [*]	NCBI RefSeq● - Nov 2016
<i>Caenorhabditis brenneri</i>	<i>Caenorhabditis sp. 10</i>	20	4	1339	106		[RGB ⁺ 14] ⁺	same
<i>Caenorhabditis elegans</i> (Orsay)	<i>C. elegans</i> ECA396 ECA723 ECA744	573	3	165	5	22	[RZL ⁺ 18]	same
<i>Chlamydia trachomatis</i>	<i>C. muridarum</i>	59	1	9924	1694		[RdSB ⁺ 18] [*]	NCBI RefSeq● - Nov 2016
<i>Ciona intestinalis A</i>	<i>C. intestinalis B</i>	20	20	1491	63	377	[RGB ⁺ 14] ⁺	same
<i>Ciona intestinalis B</i>	<i>C. intestinalis A</i>	20	20	2186	67	139	[RGB ⁺ 14] ⁺	same
<i>Clostridium difficile</i>	<i>Anaerococcus prevotii</i>	11	1	192	49		[RdSB ⁺ 18] [*]	NCBI RefSeq● - Nov 2016
<i>Corvus cornix</i>	<i>C. monedula</i>	38	20	7551159	49479	269096	[PVB ⁺ 14] [*]	[VBP ⁺ 16, PVB ⁺ 14]
<i>Coturnix japonica</i>	<i>Gallus varius</i>	20	14	5061864	87069	220450	[WZH ⁺ 18] [*]	[UKMF ⁺ 16]
<i>Culex pipiens</i>	<i>C. torrentium</i>	20	4	5442	106		[RGB ⁺ 14] ⁺	same
<i>Drosophila melanogaster</i>	<i>D. simulans</i>	196	1	4662706	151138		[LCC ⁺ 15]	[SK16]
<i>Egretta garzetta</i>	<i>Pelecanus crispus</i>	10	2	9318499	361539	10242	[LLC ⁺ 14] [*]	[ZLL ⁺ 14]
<i>Emys orbicularis</i>	<i>Trachemys scripta</i>	20	4	515	14		[RGB ⁺ 14] ⁺	same
<i>Escherichia coli</i>	<i>E. fergusonii</i>	62	1	84222	6903		NCBI RefSeq● used in [LBL ⁺ 16]	same
<i>Ficedula albicollis</i>	<i>F. hypoleuca</i>	24	2	14697230	269430	229260	[ESB ⁺ 12, KHM ⁺ 16, SMQE16, BNK ⁺ 15] [*]	[BNK ⁺ 15]
<i>Gorilla gorilla</i>	ancestral allele call from [PMSK ⁺ 13]	54	Inferred	9878547	42	569321	[PMSK ⁺ 13]	same
<i>Halictus scabiosae</i>	<i>H. simplex</i>	22	2	712	10		[RGB ⁺ 14] ⁺	[Gal16]
<i>Helicobacter pylori</i>	<i>H. felis</i>	70	1	27498	8235		[RdSB ⁺ 18] [*]	NCBI RefSeq● - Nov 2016
<i>Homo sapiens</i> (Yoruba)	ancestral allele call from [Con15]	216	Inferred	19441528	105146		[Con15]	same
<i>Klebsiella pneumoniae</i>	<i>K. varicola</i>	156	1	203601	375		[RdSB ⁺ 18] [*]	NCBI RefSeq● - Nov 2016
<i>Lepus granatensis</i>	<i>L. americanus</i>	20	2	769	31	1	[RGB ⁺ 14] ⁺	[GMFG ⁺ 13]
<i>Melitaea cinxia</i>	<i>M. didyma</i>	18		1695	101		[RGB ⁺ 14] ⁺	same
<i>Messor barbarus</i>	<i>M. structor</i>	20	8	9651	50		[RGB ⁺ 14] ⁺	[Gal16]
<i>Mycobacterium tuberculosis</i>	MYCN001 - MYCN005	33	2	7142	13	78	NCBI RefSeq● - Nov 2013	same
<i>Nipponia nippon</i>	<i>Pelecanus crispus</i>	16	2	1140694	44153	2034	[LLC ⁺ 14] [*]	[ZLL ⁺ 14]
<i>Ostrea edulis</i>	<i>O. chilensis</i>	20	4	939	28		[RGB ⁺ 14] ⁺	same
<i>Pan paniscus</i>	ancestral allele call from [PMSK ⁺ 13]	26	Inferred	6293657	63	284527	[PMSK ⁺ 13]	same
<i>Pan troglodytes elioti</i>	ancestral allele call from [PMSK ⁺ 13]	20	Inferred	10009190	44	459884	[PMSK ⁺ 13]	same
<i>Parus caeruleus</i>	<i>P. major</i>	20	2	866	51	19	[RGB ⁺ 14] ⁺	[FBRG14]
<i>Parus major</i>	<i>Cyanistes caeruleus</i>	54	2	14174305	143760	126876	[QTH ⁺ 15, LGS ⁺ 16] [*]	[MKTk16]
<i>Passer domesticus</i>	<i>P. montanus</i>	16	2	18501992	90623	633399	[RTE ⁺ 18, ETT ⁺ 17, RET ⁺ 18] [*]	[ETT ⁺ 17]
<i>Phylloscopus trochilus</i>	<i>P. tristis</i>	24	40	33401127	8605	6092936	[LLL ⁺ 17] [*]	[TKS ⁺ 17]
<i>Physa acuta</i>	<i>P. gyrina</i>	18	4	4286	176		[RGB ⁺ 14] ⁺	same
<i>Pseudomonas aeruginosa</i>	<i>P. knackmussii</i>	86	1	90258	17208		[RdSB ⁺ 18] [*]	NCBI RefSeq● - Nov 2016
<i>Sepia officinalis</i>	<i>Sepiella japonica</i>	18	2	1740	52		[RGB ⁺ 14] ⁺	[Gal16]
<i>Staphylococcus aureus</i>	<i>S. epidermis</i>	152	1	30052	8694		[RdSB ⁺ 18] [*]	NCBI RefSeq● - Nov 2016
<i>Streptococcus pneumoniae</i>	<i>S. mitis</i>	32	1	49917	2468		[RdSB ⁺ 18] [*]	NCBI RefSeq● - Nov 2016
<i>Taeniopygia guttata</i>	<i>Poephila acuticauda</i>	38	40	53263038	118506	4346767	[SLs ⁺ 15] [*]	same
<i>Zea mays</i>	<i>Tripsacum dactyloides</i>	66	1	520310	214398	1	[BMHR ⁺ 17]	same

*Core genomes were computed and aligned as [RdSB⁺18], we added to the set of a species genome one further genome from the closest species (to be able to orient the changes).

number of positions where the outgroup was diallelic (such positions are excluded, since we cannot set the ancestral allele to the outgroup allele then)

★ https://github.com/harvardinformatics/shortRead_mapping_variantCalling +SFS were extracted from transcriptomes as described in [RGB⁺14].

● NCBI RefSeq <ftp://ftp.ncbi.nih.gov/genomes/> [TCF⁺14]

○ pipeline.lbl.gov/data/araTha04_Araly1/

References

- [ABAB⁺16] Carlos Alonso-Blanco, Jorge Andrade, Claude Becker, Felix Bemm, Joy Bergelson, Karsten M. Borgwardt, Jun Cao, Eunyong Chae, Todd M. Dezwaan, Wei Ding, Joseph R. Ecker, Moises Exposito-Alonso, Ashley Farlow, Joffrey Fitz, Xiangchao Gan, Dominik G. Grimm, Angela M. Hancock, Stefan R. Henz, Svante Holm, Matthew Horton, Mike Jarsulic, Randall A. Kerstetter, Arthur Korte, Pamela Korte, Christa Lanz, Cheng-Ruei Lee, Dazhe Meng, Todd P. Michael, Richard Mott, Ni Wayan Mulyati, Thomas Nägele, Matthias Nagler, Viktoria Nizhynska, Magnus Nordborg, Polina Yu. Novikova, F. Xavier Picó, Alexander Platzter, Fernando A. Rabanal, Alex Rodriguez, Beth A. Rowan, Patrice A. Salomé, Karl J. Schmid, Robert J. Schmitz, Ümit Seren, Felice Gianluca Sperone, Mitchell Sudkamp, Hannes Svandal, Matt M. Tanzer, Donald Todd, Samuel L. Volchenbourn, Congmao Wang, George Wang, Xi Wang, Wolfram Weckwerth, Detlef Weigel, and Xuefeng Zhou. 1,135 genomes reveal the global pattern of polymorphism in *arabidopsis thaliana*. *Cell*, 166(2):481–491, 2016.
- [BMHR⁺17] Jean-Tristan Brandenburg, Tristan Mary-Huard, Guillem Rigau, Sarah J. Hearne, Hélène Corti, Johann Joets, Clémentine Vitte, Alain Charcosset, Stéphane D. Nicolas, and Maud I. Tenaillon. Independent introductions and admixtures have contributed to adaptation of european maize and its american counterparts. *PLOS Genetics*, 13(3):e1006666, Mar 2017.
- [BNK⁺15] Reto Burri, Alexander Nater, Takeshi Kawakami, Carina F Mugal, Pall I Olason, Linnea Smeds, Alexander Suh, Ludovic Dutoit, Stanislav Bureš, Laszlo Z Garamszegi, et al. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *ficedula* flycatchers. *Genome research*, 25(11):1656–1665, 2015.
- [Con15] The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*, 526(7571):68–74, 2015.
- [ESB⁺12] Hans Ellegren, Linnea Smeds, Reto Burri, Pall I Olason, Niclas Backström, Takeshi Kawakami, Axel Künstner, Hannu Mäkinen, Krystyna Nadachowska-Brzyska, Anna Qvarnström, et al. The genomic landscape of species divergence in *ficedula* flycatchers. *Nature*, 491(7426):756–760, 2012.
- [ETT⁺17] Tore O Elgvin, Cassandra N Trier, Ole K Tørresen, Ingerid J Hagen, Sigbjørn Lien, Alexander J Nederbragt, Mark Ravinet, Henrik Jensen, and Glenn-Peter Sætre. The genomic mosaicism of hybrid speciation. *Science advances*, 3(6):e1602996, 2017.
- [FBRG14] Emeric Figuet, Marion Ballenghien, Jonathan Romiguier, and Nicolas Galtier. Biased Gene Conversion and GC-Content Evolution in the Coding Sequences of Reptiles and Vertebrates. *Genome Biology and Evolution*, 7(1):240–250, 12 2014.
- [Gal16] Nicolas Galtier. Adaptive protein evolution in animals and the effective population size hypothesis. *PLOS Genetics*, 12(1):e1005774, jan 2016.
- [GMFG⁺13] Philippe Gayral, José Melo-Ferreira, Sylvain Glémin, Nicolas Bierne, Miguel Carneiro, Benoît Nabholz, Joao M. Lourenco, Paulo C. Alves, Marion Ballenghien, Nicolas Faivre, Khalid Belkhir, Vincent Cahais, Etienne Loire, Aurélien Bernard, and Nicolas Galtier. Reference-free population genomics from next-generation transcriptome data and the vertebrate–invertebrate gap. *PLOS Genetics*, 9(4):1–15, 04 2013.

- [HHW⁺17] Zachary R Hanna, James B Henderson, Jeffrey D Wall, Christopher A Emerling, Jérôme Fuchs, Charles Runckel, David P Mindell, Rauri CK Bowie, Joseph L DeRisi, and John P Dumbacher. Northern spotted owl (*strix occidentalis caurina*) genome: divergence with the barred owl (*strix varia*) and characterization of light-associated genes. *Genome biology and evolution*, 9(10):2522–2545, 2017.
- [KHM⁺16] Marty Kardos, Arild Husby, S Eryn McFarlane, Anna Qvarnström, and Hans Ellegren. Whole-genome resequencing of extreme phenotypes in collared flycatchers highlights the difficulty of detecting quantitative trait loci in natural populations. *Molecular Ecology Resources*, 16(3):727–741, 2016.
- [LBL⁺16] Marguerite Lapierre, Camille Blin, Amaury Lambert, Guillaume Achaz, and Eduardo PC Rocha. The impact of selection, gene conversion, and biased sampling on the assessment of microbial demography. *Molecular biology and evolution*, 33(7):1711–1725, 2016.
- [LCC⁺15] Justin B Lack, Charis M Cardeno, Marc W Crepeau, William Taylor, Russell B Corbett-Detig, Kristian A Stevens, Charles H Langley, and John E Pool. The *Drosophila* Genome Nexus: A Population Genomic Resource of 623 *Drosophila melanogaster* Genomes, Including 197 from a Single Ancestral Range Population. *Genetics*, 199(4):1229–1241, 01 2015.
- [LGS⁺16] Veronika N Laine, Toni I Gossmann, Kyle M Schachtschneider, Colin J Garroway, Ole Madsen, Koen JF Verhoeven, Victor De Jager, Hendrik-Jan Megens, Wesley C Warren, Patrick Minx, et al. Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nature communications*, 7(1):1–9, 2016.
- [LLC⁺14] Shengbin Li, BO Li, Cheng Cheng, Zijun Xiong, Qingbo Liu, Jianghua Lai, Hannah V Carey, Qiong Zhang, Haibo Zheng, Shuguang Wei, et al. Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species. *Genome biology*, 15(12):1–17, 2014.
- [LLL⁺17] Max Lundberg, Miriam Liedvogel, Keith Larson, Hanna Sigeman, Mats Grahn, Anthony Wright, Susanne Åkesson, and Staffan Bensch. Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. *Evolution Letters*, 1(3):155–168, 2017.
- [MKB⁺18] Jakob C Mueller, Heiner Kuhl, Stefan Boerno, Jose L Tella, Martina Carrete, and Bart Kempenaers. Evolution of genomic variation in the burrowing owl in response to recent colonization of urban areas. *Proceedings of the Royal Society B: Biological Sciences*, 285(1878):20180206, 2018.
- [MKTK16] Jakob C Mueller, Heiner Kuhl, Bernd Timmermann, and Bart Kempenaers. Characterization of the genome and transcriptome of the blue tit *cyanistes caeruleus*: polymorphisms, sex-biased expression and selection signals. *Molecular ecology resources*, 16(2):549–561, 2016.
- [PMSK⁺13] Javier Prado-Martinez, Peter H. Sudmant, Jeffrey M. Kidd, Heng Li, Joanna L. Kelley, Belen Lorente-Galdos, Krishna R. Veeramah, August E. Woerner, Timothy D. O’Connor, Gabriel Santpere, Alexander Cagan, Christoph Theunert, Ferran Casals, Hafid Laayouni, Kasper Munch, Asger Hobolth, Anders E. Halager, Maika Malig, Jessica Hernandez-Rodriguez, Irene Hernando-Herraez, Kay Prüfer, Marc Pybus, Laurel Johnstone, Michael Lachmann, Can Alkan,

- Dorina Twigg, Natalia Petit, Carl Baker, Fereydoun Hormozdiari, Marcos Fernandez-Callejo, Marc Dabad, Michael L. Wilson, Laurie Stevison, Cristina Camprubí, Tiago Carvalho, Aurora Ruiz-Herrera, Laura Vives, Marta Mele, Teresa Abello, Ivanela Kondova, Ronald E. Bontrop, Anne Pusey, Felix Lankester, John A. Kiyang, Richard A. Bergl, Elizabeth Lonsdorf, Simon Myers, Mario Ventura, Pascal Gagneux, David Comas, Hans Siegismund, Julie Blanc, Lidia Agueda-Calpena, Marta Gut, Lucinda Fulton, Sarah A. Tishkoff, James C. Mullikin, Richard K. Wilson, Ivo G. Gut, Mary Katherine Gonder, Oliver A. Ryder, Beatrice H. Hahn, Arcadi Navarro, Joshua M. Akey, Jaume Bertranpetit, David Reich, Thomas Mailund, Mikkel H. Schierup, Christina Hvilsom, Aida M. Andrés, Jeffrey D. Wall, Carlos D. Bustamante, Michael F. Hammer, Evan E. Eichler, and Tomas Marques-Bonet. Great ape genetic diversity and population history. *Nature*, 499(7459):471–475, 2013.
- [PVB⁺14] Jelmer W Poelstra, Nagarjun Vijay, Christen M Bossu, Henrik Lantz, Bettina Ryll, Inge Müller, Vittorio Baglione, Per Unneberg, Martin Wikelski, Manfred G Grabherr, et al. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, 344(6190):1410–1414, 2014.
- [QTH⁺15] Yanhua Qu, Shilin Tian, Naijian Han, Hongwei Zhao, Bin Gao, Jun Fu, Yalin Cheng, Gang Song, Per GP Ericson, Yong E Zhang, et al. Genetic responses to seasonal variation in altitudinal stress: whole-genome resequencing of great tit in eastern himalayas. *Scientific Reports*, 5(1):1–10, 2015.
- [RdSB⁺18] Olaya Rendueles, Jorge A. Moura de Sousa, Aude Bernheim, Marie Touchon, and Eduardo P. C. Rocha. Genetic exchanges are more frequent in bacteria encoding capsules. *PLOS Genetics*, 14(12):1–25, 12 2018.
- [RET⁺18] Mark Ravinet, Tore Oldeide Elgvin, Cassandra Trier, Mansour Aliabadian, Andrey Gavrilov, and Glenn-Peter Sætre. Signatures of human-commensalism in the house sparrow genome. *Proceedings of the Royal Society B*, 285(1884):20181246, 2018.
- [RGB⁺14] J. Romiguier, P. Gayral, M. Ballenghien, A. Bernard, V. Cahais, A. Chenuil, Y. Chiari, R. Dernat, L. Duret, N. Faivre, and et al. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, 515(7526):261–263, Aug 2014.
- [RTE⁺18] Anna Runemark, Cassandra N Trier, Fabrice Eroukmanoff, Jo S Hermansen, Michael Matschiner, Mark Ravinet, Tore O Elgvin, and Glenn-Peter Sætre. Variation and constraints in hybrid genome formation. *Nature Ecology & Evolution*, 2(3):549–556, 2018.
- [RZL⁺18] Aurélien Richaud, Gaotian Zhang, Daehan Lee, Junho Lee, and Marie-Anne Félix. The local coexistence pattern of selfing genotypes in *Caenorhabditis elegans* natural metapopulations. *Genetics*, 208(2):807–821, 2018.
- [SK16] Craig E. Stanley and Rob J. Kulathinal. Genomic signatures of domestication on neurogenetic genes in *Drosophila melanogaster*. *BMC Evolutionary Biology*, 16(1):6, 2016.
- [SLS⁺15] Sonal Singhal, Ellen M Leffler, Keerthi Sannareddy, Isaac Turner, Oliver Venn, Daniel M Hooper, Alva I Strand, Qiye Li, Brian Raney, Christopher N Balakrishnan, et al. Stable recombination hotspots in birds. *Science*, 350(6263):928–932, 2015.

- [SMQE16] Linnéa Smeds, Carina F Mugal, Anna Qvarnström, and Hans Ellegren. High-resolution mapping of crossover and non-crossover recombination events by whole-genome re-sequencing of an avian pedigree. *PLoS genetics*, 12(5):e1006044, 2016.
- [TCF⁺14] Tatiana Tatusova, Stacy Ciufo, Scott Federhen, Boris Fedorov, Richard McVeigh, Kathleen O'Neill, Igor Tolstoy, and Leonid Zaslavsky. Update on RefSeq microbial genomes resources. *Nucleic Acids Research*, 43(D1):D599–D605, 12 2014.
- [TKS⁺17] Venkat Talla, Faheema Kalsoom, Daria Shipilina, Irina Marova, and Niclas Backström. Heterogeneous patterns of genetic diversity and differentiation in european and siberian chiffchaff (*phylloscopus collybita abietinus*/p. *tristis*). *G3: Genes, Genomes, Genetics*, 7(12):3983–3998, 2017.
- [UKMF⁺16] Maria Ulfah, Ryouka Kawahara-Miki, Achmad Farajallah, Muladno Muladno, Ben Dorshorst, Alison Martin, and Tomohiro Kono. Genetic features of red and green junglefowls and relationship with indonesian native chickens sumatera and kedu hitam. *BMC genomics*, 17(1):1–9, 2016.
- [VBP⁺16] Nagarjun Vijay, Christen M Bossu, Jelmer W Poelstra, Matthias H Weissensteiner, Alexander Suh, Alexey P Kryukov, and Jochen BW Wolf. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature communications*, 7(1):13195, 2016.
- [WZH⁺18] Yan Wu, Yaolei Zhang, Zhuocheng Hou, Guangyi Fan, Jinsong Pi, Shuai Sun, Jiang Chen, Huaqiao Liu, Xiao Du, Jie Shen, et al. Population genomic data reveal genes related to important traits of quail. *GigaScience*, 7(5):giy049, 2018.
- [ZLL⁺14] Guojie Zhang, Cai Li, Qiye Li, Bo Li, Denis M Larkin, Chul Lee, Jay F Storz, Agostinho Antunes, Matthew J Greenwold, Robert W Meredith, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, 346(6215):1311–1320, 2014.