

Whole Genome Sequence of Two Wild-Derived *Mus musculus domesticus* Inbred Strains, LEWES/EiJ and ZALENDE/EiJ, with Different Diploid Numbers

Andrew P. Morgan,^{*,1} John P. Didion,^{*,1,2} Anthony G. Doran,[†] James M. Holt,[‡] Leonard McMillan,[‡] Thomas M. Keane,[†] and Fernando Pardo-Manuel de Villena^{*,3}

^{*}Department of Genetics, Lineberger Comprehensive Cancer Center and [‡]Department of Computer Science, University of North Carolina, Chapel Hill, North Carolina 27599-7264, [†]Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1HH, United Kingdom

ORCID IDs: 0000-0003-1942-4543 (A.P.M.); 0000-0002-8111-6261 (J.P.D.); 0000-0002-5738-5795 (F.P.-M.d.V.)

ABSTRACT Wild-derived mouse inbred strains are becoming increasingly popular for complex traits analysis, evolutionary studies, and systems genetics. Here, we report the whole-genome sequencing of two wild-derived mouse inbred strains, LEWES/EiJ and ZALENDE/EiJ, of *Mus musculus domesticus* origin. These two inbred strains were selected based on their geographic origin, karyotype, and use in ongoing research. We generated 14× and 18× coverage sequence, respectively, and discovered over 1.1 million novel variants, most of which are private to one of these strains. This report expands the number of wild-derived inbred genomes in the *Mus* genus from six to eight. The sequence variation can be accessed via an online query tool; variant calls (VCF format) and alignments (BAM format) are available for download from a dedicated ftp site. Finally, the sequencing data have also been stored in a lossless, compressed, and indexed format using the multi-string Burrows-Wheeler transform. All data can be used without restriction.

KEYWORDS inbred mouse strains; wild-derived mouse strains; Robertsonian translocations; karyotype evolution

Mus musculus (the house mouse) is among the most commonly used scientific model organisms (Didion and Pardo-Manuel de Villena 2013). “Classical” inbred strains and outbred mouse stocks used in research are primarily derived from a small founder population of *M. m. domesticus*, and therefore only sample a minor fraction of the genetic diversity present in the species as a whole (Yang *et al.* 2011; Keane *et al.* 2011). “Wild-derived” strains created by inbreeding of wild-caught mice have provided key phylogenetic context to mouse research. These relatively new strains are helping to revolutionize systems

biology by increasing the resolution of genetic mapping studies, and by expanding the range of phenotypes and disease models available to researchers (Guénet and Bonhomme 2003; Phifer-Rixey and Nachman 2015). For example, studies of chromosomal abnormalities have greatly benefited from the availability of inbred strains derived from the chromosomal races of *M. m. domesticus*, which have fixed Robertsonian (Rb) translocations involving many different autosomes (Nachman and Searle 1995; Chmátal *et al.* 2014). Wild-derived strains have also enabled phylogenetic studies that improve our understanding of the relationships between classical strains and their wild ancestors (Ideraabdullah *et al.* 2004; Yang *et al.* 2011). Recently, a few wild-derived strains have been included among the selected few parental strains used to generate new panels of consomic lines (Gregorová *et al.* 2008; Takada *et al.* 2008) and popular genetics reference populations such as the Collaborative Cross (Collaborative Cross Consortium 2012) and the Diversity Outbred (Svenson *et al.* 2012; Chick *et al.* 2016).

However, the space of wild-derived inbred strains sequenced remains limited (CAST/EiJ, PWK/PhJ, SPRET/EiJ, MOLF/EiJ, MSM/Ms, and WSB/EiJ). Most taxa, including the *M. m. domesticus* subspecies, are currently represented by a single inbred strain—a limitation compounded by the fact that many wild-derived inbred strains carry interspecific introgressions (Yang *et al.* 2011).

Copyright © 2016 Morgan *et al.*

doi: 10.1534/g3.116.034751

Manuscript received August 22, 2016; accepted for publication September 19, 2016; published Early Online October 7, 2016.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

¹These authors contributed equally to this work.

²Present address: National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892-2152.

³Corresponding author: Department of Genetics, 120 Mason Farm Road, CB#7264, University of North Carolina, Chapel Hill, NC 27599-7264. E-mail: fernando@med.unc.edu

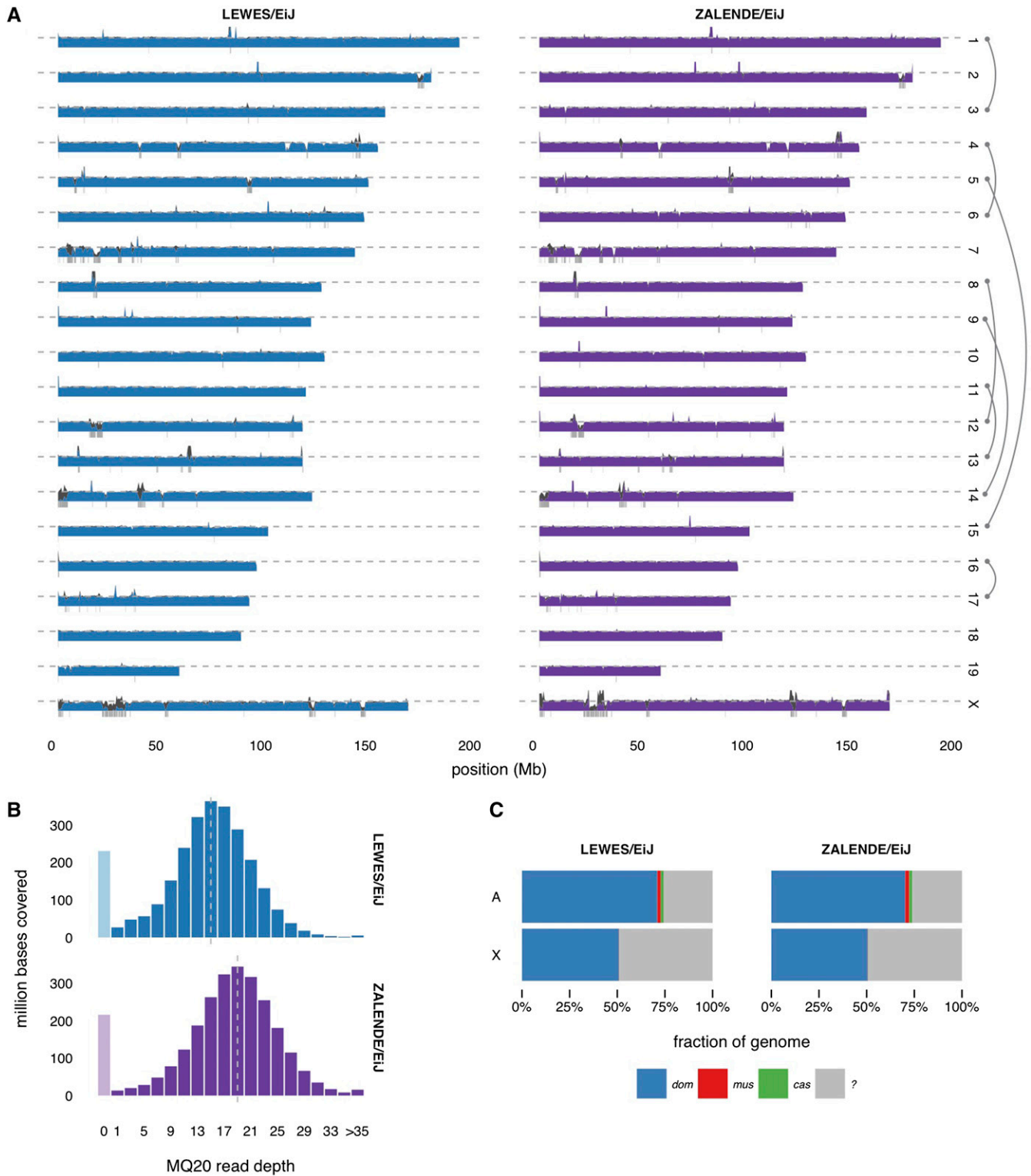


Figure 1 Sequencing coverage. (A) Profiles of normalized read depth across autosomes and X chromosome. Blue and purple regions show coverage by reads with mapping quality (MQ) > 20; dark gray regions show coverage by reads with MQ < 20. Gray dashed line indicates expected haploid depth. Boxes below axis show positions of segmental duplications > 100 kb in the mm10 reference genome. Note that the results for ZALENDE/EiJ are projections onto the reference genome, since this strain has only 13 chromosomes; chromosome pairs involved in Robertsonian fusions are indicated at right. (B) Histograms of coverage by reads with MQ > 20 across the genome. Gray dashed line indicates median coverage in each sample. (C) Estimated ancestry proportions (dom, *M. m. domesticus*; mus, *M. m. musculus*; cas, *M. m. castaneus*; ?, masked) on autosomes and X chromosome.

Here, we report the whole-genome sequencing of two *M. m. domesticus* wild-derived inbred strains: LEWES/EiJ, which is derived from mice trapped in Lewes, DE, with the standard *M. musculus* diploid chromosome number of 40; and ZALENDE/EiJ, which is derived from mice trapped in the Poschiavinus Valley (Zalende, Switzerland), and has a 26-chromosome karyotype due to fixation of seven Rb translocations.

We sequenced LEWES/EiJ and ZALENDE/EiJ to enable several specific lines of inquiry. First, this effort triples the number of *M. m. domesticus* strains sequenced, and thus provides a clearer picture of intra-subspecific variation. LEWES/EiJ and ZALENDE/EiJ were among the top 10 inbred strains recommended for resequencing based on their potential to increase the catalog of known mouse variants (Kirby *et al.* 2010).

In addition, Rb translocations have been implicated in the evolution of the mammalian karyotype, and are subject to meiotic drive during female meiosis in both mouse and humans (Pardo-Manuel de Villena and Sapienza 2001b). Our laboratory is interested in the chromosomal races of *M. m. domesticus* (Gropp and Winking 1981; Qumsiyeh 1994; Piálek *et al.* 2005; Garagna *et al.* 2014) as a model of nonrandom chromosome segregation, karyotype evolution, and the mechanisms underlying the relatively high rates of aneuploidy and trisomy in humans (Pardo-Manuel de Villena and Sapienza 2001a). We were particularly interested in whether fixation of multiple Rb translocations (*e.g.*, in ZALENDE/EiJ) is associated with losses or gains on the centromeric ends of the chromosomes involved.

LEWES/EiJ was selected because it has been used extensively in studies of male sterility in *M. m. domesticus* × *M. m. musculus* hybrids (Good *et al.* 2008). Knowledge of the specific alleles carried by LEWES/EiJ at key hybrid sterility loci will guide interpretation of those studies.

Finally, we recently discovered a large copy number variant on Chromosome 2, *R2d2*, for which the high copy number allele is associated with distorted transmission ratios in heterozygous female carriers (Didion *et al.* 2015). Importantly, we found that *R2d2* has driven selective sweeps in the absence of fitness gain (“selfish sweeps”) in multiple independent mouse populations (Didion *et al.* 2016). *R2d2* was first identified in WSB/EiJ, a *M. m. domesticus* strain derived from mice trapped in Centreville, MD. SNP genotyping data (Yang *et al.* 2011; Didion *et al.* 2012; Morgan *et al.* 2016a) indicated that ZALENDE/EiJ harbors the high-copy (*i.e.*, distortion-associated) *R2d2* allele, while LEWES/EiJ—derived from mice trapped only ~100 km away from Centreville—has the low-copy (*i.e.*, wild-type) *R2d2* allele. We hypothesized that comparison of the WSB/EiJ genome, which was sequenced previously (Keane *et al.* 2011), to the LEWES/EiJ and ZALENDE/EiJ genomes would help to fine-map *R2d2*, and characterize its evolutionary history (Morgan *et al.* 2016b), and could be of further use in understanding the relationship between copy number and transmission distortion at this locus.

Here, we describe the basic characteristics of these two genomes, make them available for public use, and discuss how this resource may benefit the community.

MATERIALS AND METHODS

Mouse strains

LEWES/EiJ (<https://www.jax.org/strain/002798>): Derived from wild mice trapped in Lewes, DE. Mice were sent from Michael Potter (National Cancer Institute) to Eva M. Eicher at The Jackson Laboratory in 1996.

ZALENDE/EiJ (<https://www.jax.org/strain/001392>): Derived from mice trapped in the Poschiavinus Valley (Zalende, Switzerland) by Richard D. Sage. Mice from Sage’s colony were transferred to Michael Potter in 1981. A single pair of mice was sent from Michael Potter to

■ **Table 1 Variant-calling statistics**

| | LEWES/EiJ | | ZALENDE/EiJ | | Shared | |
|----------------|-----------|-------|-------------|-------|---------|-------|
| Private SNVs | 403,770 | | 612,721 | | 102,561 | |
| Coding | 3619 | 0.90% | 5525 | 0.90% | 672 | 0.66% |
| Damaging | 39 | 0.01% | 67 | 0.01% | 3 | 0.00% |
| Private indels | 92,082 | | 157,366 | | 19,684 | |
| Coding | 51 | 0.06% | 86 | 0.05% | 9 | 0.05% |
| Damaging | 49 | 0.05% | 77 | 0.05% | 7 | 0.04% |

Eva M. Eicher at The Jackson Laboratory in 1982. ZALENDE/EiJ is homozygous for seven Rb translocations (1.3, 4.6, 5.15, 11.13, 8.12, 9.14, and 16.17).

Sequencing

High molecular weight DNA from one ZALENDE/EiJ male was obtained from The Jackson Laboratory. High molecular weight DNA from a LEWES/EiJ female was prepared from tissues samples from a colony maintained in the Pardo-Manuel de Villena laboratory in the past (Bell *et al.* 2006)

Libraries were prepared and sequenced at the University of North Carolina High Throughput Sequencing Facility. Genomic DNAs were sheared by ultrasonication and the resulting fragments were size-selected to target size 350 bp using a PippinPrep system. Samples were barcoded (LEWES/EiJ, two barcodes; ZALENDE/EiJ, four barcodes), pooled, and sequenced across multiple lanes and multiple flowcells on an Illumina HiSeq 2000 instrument. After a small pilot run with single-end 50-bp reads, and a mixture of single- and paired-end 100-bp reads were generated for each sample. Base calling and demultiplexing were performed using the Casava 1.8 pipeline. We obtained 498,668,400 reads for LEWES/EiJ and 573,861,165 reads for ZALENDE/EiJ.

Data processing

Integrity of raw sequencing reads was confirmed using FastQC (Andrews 2010). Reads were aligned to the mm10 reference genome using bwa mem v0.7.5a-r406 (Li 2013). Coverage and quality summaries were computed using the Picard suite (<http://broadinstitute.github.io/picard>).

SNV and short indel variants were called using the Sanger Mouse Genomes Project pipeline, the current version of which is described in detail elsewhere (Doran *et al.* 2016). Briefly, samtools mpileup v1.1 and bcftools call v1.1 were used to generate an initial call set for the nuclear genome and mitochondrial genome. Candidate variants were filtered on the basis of read depth (> 5, < 100 for nuclear genome; > 350 for mitochondrial genome), mapping quality (> 20), number of reads supporting the alternate allele (> 5), proximity to an indel (> 2 bp; SNVs only) and homozygosity. Variants were declared private to a strain if the alternate allele was absent from all 30 other strains in the Mouse Genomes Project catalog. Analyses presented in this manuscript were performed only on variants passing all filters.

Burrows-Wheeler transforms

The multi-string Burrows-Wheeler transform (BWT) for each whole genome dataset was individually constructed in-memory using the ropebwt2 program (Li 2014). After construction, each BWT was encoded using the run-length encoding format of the msbwt program (Holt and McMillan 2014) for disk storage. The BWTs for LEWES/EiJ and ZALENDE/EiJ are 9.6 and 10.1 GB in size, respectively. Instructions for building BWTs are publicly available at <https://github.com/holtjma/msbwt/wiki/Converting-to-msbwt's-RLE-format>.

■ Table 2 Deletions of protein-coding genes

| Strain | Ensembl Gene ID | Gene Symbol | Chromosome |
|--------------------|--------------------|----------------------|------------|
| LEWES/EiJ | ENSMUSG00000070868 | <i>Skint3</i> | 4 |
| | ENSMUSG00000055960 | <i>Skint4</i> | 4 |
| ZALENDE/EiJ | ENSMUSG00000073609 | <i>D2hgdh</i> | 1 |
| | ENSMUSG00000094651 | <i>Gal3st2</i> | 1 |
| | ENSMUSG00000093805 | <i>Gm9994</i> | 1 |
| | ENSMUSG00000089951 | <i>Gm14435</i> | 2 |
| | ENSMUSG00000070868 | <i>Skint3</i> | 4 |
| | ENSMUSG00000055960 | <i>Skint4</i> | 4 |
| | ENSMUSG00000049972 | <i>Skint9</i> | 4 |
| | ENSMUSG00000055594 | <i>5530400C23Rik</i> | 6 |
| | ENSMUSG00000067599 | <i>Klra7</i> | 6 |
| | ENSMUSG00000091620 | <i>Vmn2r23</i> | 6 |
| | ENSMUSG00000094298 | <i>Gm6164</i> | 7 |
| | ENSMUSG00000094981 | <i>Gm8653</i> | 7 |
| | ENSMUSG00000093941 | <i>Vmn1r131</i> | 7 |
| | ENSMUSG00000091195 | <i>Gm17332</i> | 11 |
| | ENSMUSG00000091275 | <i>Gm3248</i> | 14 |
| | ENSMUSG00000096345 | <i>Esp16</i> | 17 |
| | ENSMUSG00000079342 | <i>Lipo1</i> | 19 |
| ENSMUSG00000079387 | <i>Luzp4</i> | X | |

Data availability

Raw reads have been deposited in the European Nucleotide Archive (accession #PRJEB15190). All processed data are available from the Sanger Mouse Genomes Project (<http://www.sanger.ac.uk/science/data/mouse-genomes-project>). Aligned reads (in BAM format): ftp://ftp-mouse.sanger.ac.uk/current_bams; web interface for querying variants: http://www.sanger.ac.uk/sanger/Mouse_SnpViewer/rel-1505; bulk download of variants (VCF format): ftp://ftp-mouse.sanger.ac.uk/current_snps.

BWTs are available for download at <http://csbio.unc.edu/WildDerived>.

RESULTS AND DISCUSSION

We sequenced one female LEWES/EiJ individual to median 14× coverage (overall alignment rate 99.5%), and one male ZALENDE/EiJ individual to median 18× coverage (alignment rate 99.4%). Genetic sex was confirmed by comparing relative read depth on the X chromosome to the autosomes. Coverage profiles across the nuclear genome are

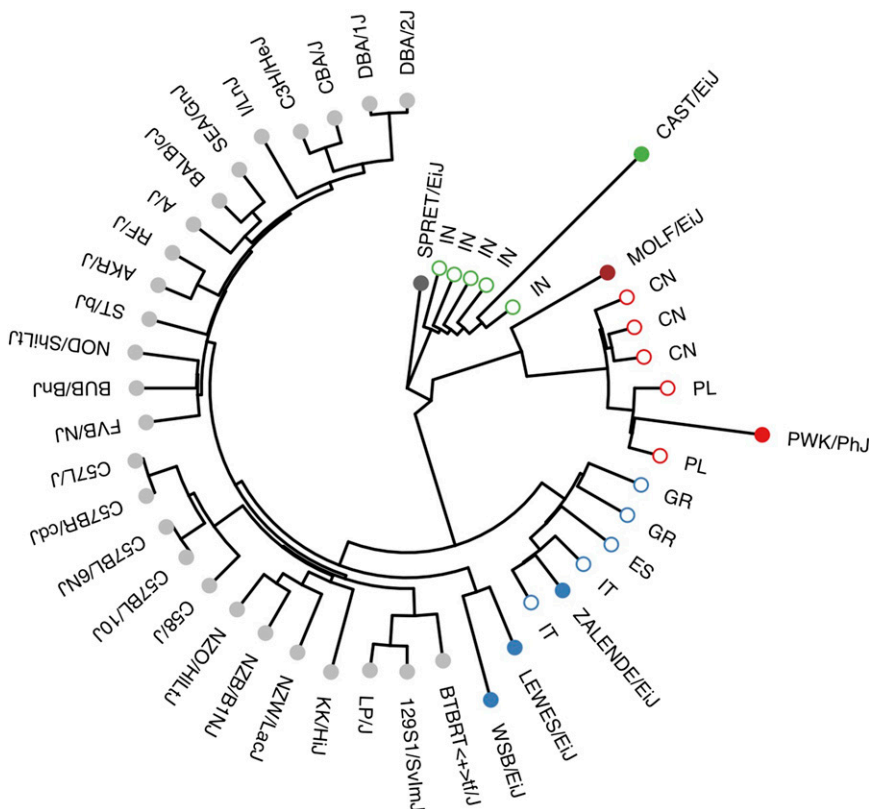


Figure 2 Phylogenetic tree of Mouse Genomes Project strains. Tree was constructed from genotypes at 30,000 ancestry-informative SNPs from the Mouse Diversity Array identified in Yang et al. (2011). Filled dots, inbred strains sequenced by Sanger MGP; open dots, wild-caught mice from Yang et al. (2011), identified by two-letter country code. Samples are colored according to subspecies of origin: blue, *M. m. domesticus*; red, *M. m. musculus*; maroon, *M. m. molossinus* (MOLF/EiJ); green, *M. m. castaneus*. The tree was rooted using SPRET/EiJ (*M. spretus*) as the outgroup.

shown in Figure 1A; corresponding histograms appear in Figure 1B. After filtering reads with ambiguous alignments or poor base-call quality, coverage of at least 10× was achieved over 77.2 and 84.7% of the genome in each sample, respectively. This represents the fraction of the genome accessible for identification of sequence variants. The remaining fraction of the genome lies almost entirely in repetitive elements and clusters of polymorphic segmental duplications (e.g., the proximal regions of chromosomes 7 and 14) where unambiguous alignment of short reads is not possible.

Sequence variants (SNVs and short indels < 10 bp in size) were ascertained using the Sanger Mouse Genomes Project pipeline (Doran *et al.* 2016). We identified 1,119,052 SNVs that have not been previously reported in any other mouse inbred strain (Keane *et al.* 2011; Doran *et al.* 2016). Of those variants, 102,561 are shared exclusively by the two strains, while 403,770 and 612,721 SNVs are private to LEWES/Eij and ZALENDE/Eij, respectively (Table 1), with a transition:transversion ratio of 2.14. Comparison of these totals with the number of unique variants discovered in other inbred strains reveals that sequencing of wild-derived inbred strains of *M. m. domesticus* origin identifies at least one order of magnitude more variants than sequencing classical laboratory strains (Table 2) (Keane *et al.* 2011; Doran *et al.* 2016).

Of the 1.1 M new variants, 0.88% fall within coding sequences and three (shared), 39 (private to LEWES/Eij), and 67 (private to ZALENDE/Eij) are predicted to disrupt gene function. We observe a similar picture for small indels (Table 1). As expected, the number of small indels that fall with coding exons is smaller but the proportion of predicted damaging mutations is higher.

We also identified large deletions that are predicted to encompass multiple exons of at least 20 genes (Table 2). These deletions can be ascribed to 11 events, and represent natural knockouts that are compatible with life in a laboratory setting. Most of them affect members of large and highly polymorphic gene families. Those that affect single-copy genes of well-defined function such as *D2hgdh* and *Gal3st2*, are also present in other sequenced strains (Keane *et al.* 2011, Doran *et al.* 2016). Interestingly, both the number of deletion events, and genes deleted, appears to be higher in ZALENDE/Eij than in LEWES/Eij. We speculate that differences between the two strains in both their phylogenetic proximity to the C57BL/6J reference sequence, and the effective population sizes of their wild progenitors, may explain these differences.

Both LEWES/Eij and ZALENDE/Eij were reported to have essentially pure *M. m. domesticus* ancestry based on genotypes from the 600 K SNP Mouse Diversity Array (Yang *et al.* 2011; Figure 2). We sought to confirm this result using dense genotypes from whole-genome sequencing. We used three wild-derived strains already sequenced by the Mouse Genomes Project—WSB/Eij (*M. m. domesticus*), PWK/PhJ (*M. m. musculus*) and CAST/Eij (*M. m. castaneus*)—as representatives for their subspecies, and SPRET/Eij (*Mus spretus*) as an outgroup. We classified alleles as ancestral or derived based on the pattern of sharing with the outgroup. After masking the 19% of the genome with known intersubspecific introgression or contamination in these strains (Yang *et al.* 2011), we calculated (in 25-kb windows) the proportion of derived alleles shared between LEWES/Eij and exactly one of WSB/Eij, PWK/PhJ, or CAST/Eij, and repeated this analysis with ZALENDE/Eij. Both strains share a majority of derived alleles with WSB/Eij (*M. m. domesticus*) over 92 and 94% of the genome, respectively (Figure 1C). The remaining 8 and 6% of windows represent either introgression from *M. m. musculus* or *M. m. castaneus* in the two wild-derived inbred strains sequenced here; small *M. m. domesticus* introgressions into PWK/PhJ and CAST/Eij (Wang *et al.* 2012); regions of incomplete sorting of ancestral polymorphism among the three

subspecies (Keane *et al.* 2011); homoplasy (recurrent mutation); or some combination of the these.

Finally, we investigated whether the Robertsonian translocations in ZALENDE/Eij are associated with large deletions or duplications near the centromeres of the affected chromosomes (1, 3, 4, 5, 6, 8, 9, 11, 12, 13, 14, 15, 16, and 17). We found no evidence for private CNVs in any of these centromere-proximal regions (Figure 1A). Based on the results obtained in ZALENDE/Eij, it would appear that the emergence of Rb races is not associated with large-scale changes in sequence content, at least in the regions of the genome included in the reference assembly.

Novel alleles presented here have already been used to increase the utility of SNP arrays for genotyping of wild mice (Morgan *et al.* 2016a). We provide tools to browse the variants, the underlying read alignments, and the raw reads (see *Data Availability* in *Materials and Methods*). In particular, we provide access to the BWT of the sequencing reads generated from each of these strains. A BWT is a compact and lossless data structure that allows rapid interrogation of the reads in the absence of alignment.

Integration of these two novel genome sequences with the growing catalog of known *M. musculus* genetic variation will provide a valuable resource to researchers using mouse models to study a wide variety of biological processes, including karyotype evolution (Chmátal *et al.* 2014), speciation (Payseur and Place 2007), protein evolution (Karn *et al.* 2008; Stempel *et al.* 2016), gene duplications (Laukaitis *et al.* 2008; Morgan *et al.* 2016b), X-chromosome inactivation (Calaway *et al.* 2013), and disease-associated quantitative traits (Cervino *et al.* 2005; Perez *et al.* 2013; Nicod *et al.* 2016).

ACKNOWLEDGMENTS

We wish to thank Timothy A. Bell for help preparing the DNA. This work was supported in part by National Institutes of Health (NIH) grants F30MH103925 (A.P.M.), T32GM067553 (A.P.M. and J.P.D.), P50GM076468, and U19AI100625 (F.P.-M.dV.). T.M.K. and A.G.D. are supported by the Medical Research Council (MR/L007428/1), the Biotechnology and Biological Sciences Research Council (BB/M000281/1), and the Wellcome Trust.

LITERATURE CITED

- Andrews, S., 2010 FastQC: A Quality Control Tool for High Throughput Sequence Data. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. Accessed: August 1, 2016.
- Bell, T. A., E. de la Casa-Esperón, H. E. Doherty, F. Ideraabdullah, K. Kim *et al.*, 2006 The paternal gene of the DDK syndrome maps to the Schlafen gene cluster on mouse chromosome 11. *Genetics* 172: 411–423.
- Calaway, J. D., A. B. Lenarcic, J. P. Didion, J. R. Wang, J. B. Searle *et al.*, 2013 Genetic architecture of skewed X inactivation in the laboratory mouse. *PLoS Genet.* 9: e1003853.
- Cervino, A. C., G. Li, S. Edwards, J. Zhu, C. Laurie *et al.*, 2005 Integrating QTL and high-density SNP analyses in mice to identify *Insig2* as a susceptibility gene for plasma cholesterol levels. *Genomics* 86: 505–517.
- Chick, J. M., S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi *et al.*, 2016 Defining the consequences of genetic variation on a proteome-wide scale. *Nature* 534: 500–505.
- Chmátal, L., S. I. Gabriel, G. P. Mitsainas, J. Martínez-Vargas, J. Ventura *et al.*, 2014 Centromere strength provides the cell biological basis for meiotic drive and karyotype evolution in mice. *Curr. Biol.* 24: 2295–2300.
- Collaborative Cross Consortium, 2012 The genome architecture of the collaborative cross mouse genetic reference population. *Genetics* 190: 389–401.
- Didion, J. P., and F. Pardo-Manuel de Villena, 2013 Deconstructing *Mus gemischus*: advances in understanding ancestry, structure, and variation in the genome of the laboratory mouse. *Mamm. Genome* 24: 1–20.

- Didion, J. P., H. Yang, K. Sheppard, C.-P. Fu, L. McMillan *et al.*, 2012 Discovery of novel variants in genotyping arrays improves genotype retention and reduces ascertainment bias. *BMC Genomics* 13: 34
- Didion, J. P., A. P. Morgan, A. M. F. Clayshulte, R. C. McMullan, L. Yadgary *et al.*, 2015 A multi-megabase copy number gain causes maternal transmission ratio distortion on mouse chromosome 2. *PLoS Genet.* 11: e1004850.
- Didion, J. P., A. P. Morgan, L. Yadgary, T. A. Bell, R. C. McMullan *et al.*, 2016 R2d2 drives selfish sweeps in the house mouse. *Mol. Biol. Evol.* 33: 1381–1395.
- Doran, A. G., K. Wong, J. Flint, D. J. Adams, K. W. Hunter *et al.*, 2016 Deep genome sequencing and variation analysis of 13 inbred mouse strains defines candidate phenotypic alleles, private variation and homozygous truncating mutations. *Genome Biol.* 17: 167.
- Garagna, S., J. Page, R. Fernandez-Donoso, M. Zuccotti, and J. B. Searle, 2014 The Robertsonian phenomenon in the house mouse: mutation, meiosis and speciation. *Chromosoma* 123: 529–544.
- Good, J. M., M. A. Handel, and M. W. Nachman, 2008 Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. *Evolution* 62: 50–65.
- Gregorová, S., P. Divina, R. Storchova, Z. Trachtulec, V. Fotopulosova *et al.*, 2008 Mouse consomic strains: exploiting genetic divergence between *Mus m. musculus* and *Mus m. domesticus* subspecies. *Genome Res.* 18: 509–515.
- Gropp, A., and H. Winking, 1981 Robertsonian translocations: cytology, meiosis, segregation patterns and biological consequences. *Nature* 405: 212–219.
- Guénet, J. L., and F. Bonhomme, 2003 Wild mice: an ever-increasing contribution to a popular mammalian model. *Trends Genet.* 19: 24–31.
- Holt, J., and L. McMillan, 2014 Merging of multi-string BWTs with applications. *Bioinformatics* 30: 3524–3531.
- Ideraabdullah, F. Y., E. de la Casa-Esperón, T. A. Bell, D. A. Detwiler, T. Magnuson *et al.*, 2004 Genetic and haplotype diversity among wild-derived mouse inbred strains. *Genome Res.* 14: 1880–1887.
- Karn, R. C., N. L. Clark, E. D. Nguyen, and W. J. Swanson, 2008 Adaptive evolution in rodent seminal vesicle secretion proteins. *Mol. Biol. Evol.* 25: 2301–2310.
- Keane, T. M., L. Goodstadt, P. Danecek, M. A. White, K. Wong *et al.*, 2011 Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477: 289–294.
- Kirby, A., H. M. Kang, C. M. Wade, C. Cotsapas, E. Kostem *et al.*, 2010 Fine mapping in 94 inbred mouse strains using a high-density haplotype resource. *Genetics* 185: 1081–1095.
- Laukaitis, C. M., A. Heger, T. D. Blakley, P. Munclinger, C. P. Ponting *et al.*, 2008 Rapid bursts of androgen-binding protein (Abp) gene duplication occurred independently in diverse mammals. *BMC Evol. Biol.* 8: 46.
- Li, H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arxiv*: 1303.3997
- Li, H., 2014 Fast construction of FM-index for long sequence reads. *Bioinformatics* 30: 3274–3275.
- Morgan, A. P., C.-P. Fu, C.-Y. Kao, C. E. Welsh, J. P. Didion *et al.*, 2016a The mouse universal genotyping array: from substrains to subspecies. *G3 (Bethesda)* 6: 263–279.
- Morgan, A. P., J. M. Holt, R. C. McMullan, T. A. Bell, A. M.-F. Clayshulte *et al.*, 2016b The evolutionary fates of a large segmental duplication in mouse. *Genetics* 204: 267–285.
- Nachman, M. W., and J. B. Searle, 1995 Why is the house mouse karyotype so variable? *Trends Ecol. Evol. (Amst.)* 10: 397–402.
- Nicod, J., R. W. Davies, N. Cai, C. Hassett, L. Goodstadt *et al.*, 2016 Genome-wide association of multiple complex traits in outbred mice by ultra-low-coverage sequencing. *Nat. Genet.* 48: 912–918.
- Pardo-Manuel de Villena, F., and C. Sapienza, 2001a Female meiosis drives karyotypic evolution in mammals. *Genetics* 159: 1179–1189.
- Pardo-Manuel de Villena, F., and C. Sapienza, 2001b Transmission ratio distortion in offspring of heterozygous female carriers of Robertsonian translocations. *Hum. Genet.* 108: 31–36.
- Payseur, B. A., and M. Place, 2007 Searching the genomes of inbred mouse strains for incompatibilities that reproductively isolate their wild relatives. *J. Hered.* 98: 115–122.
- Perez, C. J., A. Dumas, L. Vallières, J.-L. Guénet, and F. Benavides, 2013 Several classical mouse inbred strains, including DBA/2, NOD/Lt, FVB/N, and SJL/J, carry a putative loss-of-function allele of Gpr84. *J. Hered.* 104: 565–571.
- Phifer-Rixey, M., and M. W. Nachman, 2015 Insights into mammalian biology from the wild house mouse *Mus musculus*. *eLife* 4: e05959.
- Piálek, J., H. C. Hauffe, and J. B. Searle, 2005 Chromosomal variation in the house mouse. *Biol. J. Linn. Soc. Lond.* 84: 535–563.
- Qumsiyeh, M. B., 1994 Evolution of number and morphology of heterozygosity. *Symp. Zool. Soc. Lond.* 47: 141–181.
- Stempel, H., M. Jung, A. Pérez-Gómez, T. Leinders-Zufall, F. Zufall *et al.*, 2016 Strain-specific loss of formyl peptide receptor 3 in the murine vomeronasal and immune systems. *J. Biol. Chem.* 291: 9762–9775.
- Svenson, K. L., D. M. Gatti, W. Valdar, C. E. Welsh, R. Cheng *et al.*, 2012 High-resolution genetic mapping using the mouse diversity outbred population. *Genetics* 190: 437–447.
- Takada, T., A. Mita, A. Maeno, T. Sakai, H. Shitara *et al.*, 2008 Mouse inter-subspecific consomic strains for genetic dissection of quantitative complex traits. *Genome Res.* 18: 500–508.
- Wang, J. R., F. Pardo-Manuel de Villena, H. A. Lawson, J. M. Cheverud, G. A. Churchill *et al.*, 2012 Imputation of SNPs in inbred mice using local phylogeny. *Genetics* 190: 449.
- Yang, H., J. R. Wang, J. P. Didion, R. J. Buus, T. A. Bell *et al.*, 2011 Subspecific origin and haplotype diversity in the laboratory mouse. *Nat. Genet.* 43: 648–655.

Communicating editor: B. J. Andrews