Draft genome assemblies using sequencing reads from Oxford Nanopore Technology and Illumina platforms for four species of North American killifish from the *Fundulus* genus

Lisa K. Johnson [1,2], Ruta Sahasrabudhe [3], Tony Gill [1], Jennifer Roach [1], Lutz Froenicke [3], C. Titus Brown [2], Andrew Whitehead* [1]

[1] Department of Environmental Toxicology, University of California, Davis
[2] Department of Population Health & Reproduction, School of Veterinary Medicine, University of California, Davis
[3] DNA Technologies Core, Genome Center, University of California, Davis
*corresponding author: awhitehead@ucdavis.edu

Abstract

Draft *de novo* reference genome assemblies were obtained from four North American killifish species (*Fundulus xenicus*, *Fundulus catenatus*, *Fundulus nottii*, and *Fundulus olivaceus*) using sequence reads from Illumina and Oxford Nanopore Technologies' PromethION platforms. For each species, the PromethION platform was used to generate 30-45x sequence coverage, and the Illumina platform was used to generate 50-160x sequence coverage. Contig N50 values ranged from 0.4 Mb to 2.7 Mb, and BUSCO scores were consistently above 90% complete using the Eukaryota database. Draft assemblies and raw sequencing data are available for public use. We encourage use and re-use of these data for assembly benchmarking and external analyses.

Keywords: long reads; Oxford Nanopore; killifish; genomes; genome assembly

Background

Sequencing and assembling large eukaryotic genomes is challenging [1–3]. Accuracy of downstream analyses, such as variant calling and measuring gene expression, depends heavily on a high-quality reference genome [4]. Fortunately, the cost of generating whole genome sequence data is dropping, making it easier for individual labs rather than large consortiums to generate assemblies for organisms without reference genomes [3,5,6]. Single-molecule long read nucleic acid sequencing technology from Oxford Nanopore Technologies (ONT), which has been commercially available since 2014 [7], has been shown to improve the contiguity of reference assemblies [8] and reveal "dark regions" that were previously camouflaging genes [9]. The lengths of the sequencing reads generated using this technology are limited only by the size of the fragments in the extracted DNA sample [10]. The promise of more complete reference assemblies is especially important for the accuracy of comparative evolutionary genomics studies, as assembly fragments lead to errors in downstream synteny analyses [11], as well as

SNP calling and identification of transcript features (splice junctions and exons) for quantification.

Despite high error rates ~5% [12] relative to Illumina short reads ~0.25% [13] and the relatively recent availability of ONT data, there have been a flurry of studies using this sequencing technology. Small genomes from bacteria and viruses appear to be ideal for sequencing on the ONT MinION platform [12]. The portable nature of the technology makes it appealing as a resource for teaching [14,15], working in remote locations [16–18] and for investigating viral outbreak public health emergencies [19–21]. However, despite the demonstrated ability to achieve yields >6.5 Gb per flow cell [22], the MinION platform can be prohibitively expensive for sequencing larger eukaryotic genomes. For example, 39 flow cells yielded 91.2 Gb of sequence data (~30x coverage) of the human genome [23]. Sequencing of the wild tomato species, *Solanum pennellii* across thirty-one flow cells yielded 110.96 Gb (~100x coverage) with some flow cells yielding >5Gb [24]. By contrast, following the 2018 beta release of the ONT PromethION platform, which has a higher density of nanopore channels, five flow cells were used to yield >250 Gb (~80x coverage) of the human genome [25]. PromethION data combined with Hi-C long-range mapping data from human samples produced a genome genome assembly with a scaffold N50 of 56.4 Mbp [26].

The combination of long read sequencing data from ONT MinION and short read sequencing data from Illumina has been used to improve the quality of reference genomes [27–30]. In one approach, short read assembly scaffolds have been improved with the addition of long reads. The Murray cod genome (640-669 Mb in size) was improved by combining low coverage 804 Mb of long reads ONT data from just one MinION flow cell with 70.6 Gb of Illumina data from both HiSeq and MiSeq; the assembly scaffold N50 increased from 33,442 bp (Illumina only) to 52,687 bp with ONT and Illumina combined [31]. The clownfish genome (791 to 794 Mb in size) was improved by including 8.95 Gb of ONT MinION reads; the scaffold N50 increased from 21,802 bp (Illumina only) to 401,715 bp with ONT and Illumina combined [27]. Recently, a new approach is available with racon [32] and/or pilon [33] consensus building tools, which uses Illumina data to “polish” contigs from ONT-only assemblies. Polishing corrects single nucleotide base differences, fills gaps, and identifies local mis-assemblies [33]. This approach has been shown to improve the BUSCO score from <1% with the ONT assembly alone to >95% complete after polishing with Illumina reads [28].

In this study, we explored whether the ONT PromethION sequencing technology could be appropriate for generating initial draft reference genomes for four species of North American killifish belonging to the *Fundulus* genus. *Fundulus* is a comparative evolutionary model system for studying repeated genomic divergence between marine and freshwater species. *Fundulus* killifish have a cosmopolitan geographic distribution across North America. These small cyprinodontiform fish have evolved to occupy a wide range of osmotic niches, including marine,

estuarine, and freshwater [34]. Estuarine and coastal *Fundulus* are euryhaline, insofar as they can adjust their physiologies to tolerate a very wide range of salinities. In contrast, freshwater species are stenohaline: they tolerate a much narrower range of salinities [34,35]. Freshwater clades are derived from marine clades, and radiation into freshwater has occurred multiple times independently within the genus. This makes *Fundulus* unusual, because most large clades of fishes are either exclusively marine or exclusively freshwater. Therefore, species of closely-related killifish in the *Fundulus* genus serve as a unique comparative model system for understanding the genomic mechanisms that contribute to evolutionary divergence and convergence of osmoregulatory processes, which is important for understanding how species will evolve to face fluctuating salinities in future climate change scenarios [36]. The Atlantic killifish, *Fundulus heteroclitus* has been a well-described physiological model organism for investigating the functional basis of, and evolution of, physiological resilience to temperature, salinity, hypoxia, and environmental pollution [34,37–39], with an available genome from the Atlantic killifish, *Fundulus heteroclitus* [40]. However, we do not currently have any genomes from other *Fundulus* killifish, particularly from those occupying freshwater habitats.

Here, we report the collection of whole genome sequencing data using both ONT PromethION and Illumina platforms from four killifish species without previously-existing sequencing data (Figure 1): *Fundulus xenicus* (formerly *Adinia xenica*) [41], *Fundulus catenatus*, *Fundulus nottii*, and *Fundulus olivaceus*. *F. xenicus* is euryhaline and occupies coastal and estuarine habitats, while the other species (*F. catenatus*, *F. nottii*, *F. olivaceus*) are stenohaline and occupy freshwater habitats.

Figure 1. Four *Fundulus* killifish (left to right): the marine diamond killifish *Fundulus xenicus*; the northern studfish, *Fundulus catenatus* (south central United States); the freshwater bayou topminnow, *Fundulus nottii*; and the freshwater blackspotted topminnow, *Fundulus olivaceus*. (drawings used with permission from the artist, Joseph R. Tomelleri).

Methods and Results

Live field-caught individuals of each fish species were shipped to UC Davis and kept at their native salinities in an animal holding facility, maintained according to University of California IACUC standards. *F. catenatus* and *F. olivaceus* were collected from the Gasconade River, MO (latitude/longitude coordinates 37.879/-91.795 and 37.19/-92.56, respectively), *F. nottii* was collected from Walls Creek, MS (31.154433/-89.245381), and *F. xenicus* was collected from Graveline Bayou, MS (30.368756/-88.719329). High molecular weight (hmw) DNA was extracted from fresh tissue for *F. nottii* and *F. xenicus*, and from frozen tissue for *F. catenatus*

and *F. olivaceus*. For *F. catenatus* and *F. olivaceus*, tissues were dissected and frozen in liquid nitrogen and stored at -80 °C until samples were prepared for hmw DNA extraction. With the exception of *F. olivaceus*, each assembly consisted of sequencing one tissue sample from one individual. For *F. olivaceus*, Illumina data were collected from DNA extracted from one individual while the ONT PromethION data were collected from another individual (frozen tissue).

DNA extractions

Whole fish heads were used for hmw DNA extractions. Agilent's Genomic DNA Isolation kit (Catalog #200600) was used to extract DNA from fresh tissues from *F. xenicus* and *F. nottii*. For *F. catenatus* and *F. olivaceus*, both the ultra-long read sequencing protocol from [42] (which included Tissue lysis buffer with Tris, NaCl, EDTA, SDS and Proteinase K followed by phenol:chloroform extraction), as well as the Qiagen "DNA purification from tissue using the Gentra puregene Tissue Kit" (p. 39) were used, and were found to be similar to the Agilent kit. Precipitated DNA was difficult to re-dissolve; therefore, additional phenol:chloroform cleanup steps were required after extractions. We found that adding urea to the lysis buffer helped the precipitated DNA pellet to be less fragile and go into solution easier [43]. Prior to library preparation, hmw DNA from *F. nottii* and *F. olivaceus* (PromethION) was sheared to 50 kb in an effort to improve the ligation enzyme efficiency, resulting in fragments in the 50-70 kb range. Field inversion gels were used to visualize hmw DNA (Figure 2).

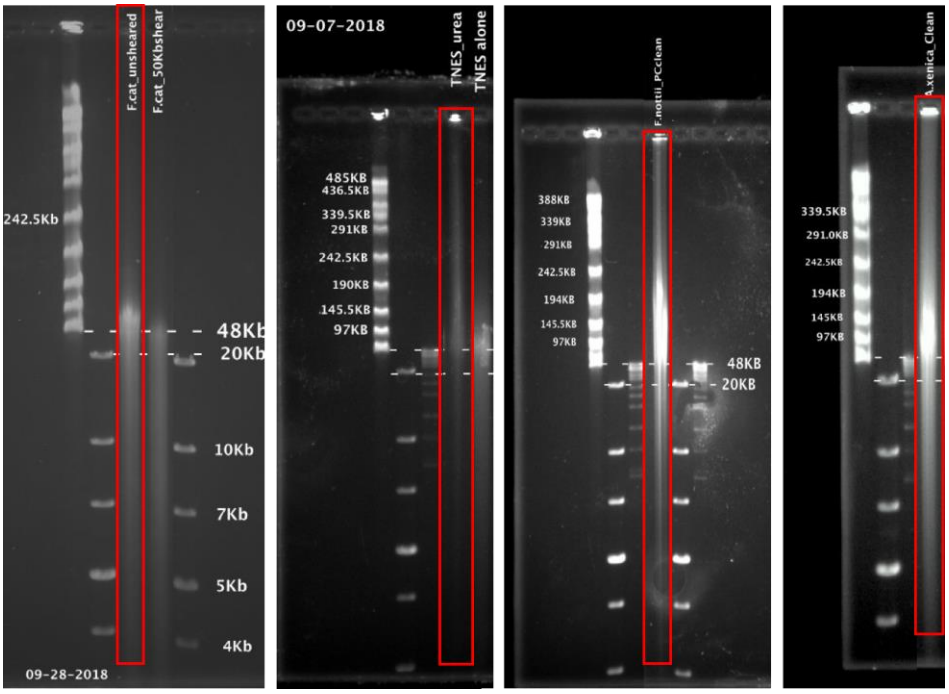

Figure 2. Field inversion gels with red boxes showing samples sequenced (in order from left to right: *F. catenatus* (sheared vs. unsheared), *F. olivaceus*, *F. nottii*, *F. xenicus*). DNA was

extracted from fresh tissues for *F. xenicus* and *F. nottii*, and from frozen tissue for *F. catenatus* and *F. olivaceus*.

ONT sequencing

Libraries for ONT PromethION sequencing were prepared using the ligation sequencing kit (SQK-LSK109) following the manufacturer's instructions. ONT PromethION sequencing data were collected from all four species on an alpha-beta instrument through the early release program at the University of California, Davis DNA Technologies Core facility (Davis, CA USA). One species was sequenced per flow cell. Base-calling was done onboard the PromethION instrument (Oxford Nanopore Technologies, UK). For the *F. xenicus* run, lambda phage (DNA CS) was spiked-in as a positive control.

Illumina Sequencing

With the exception of *F. olivaceus*, each individual hmw DNA sample used for the ONT library was also used for Illumina library preparation using the Nextera Index Kit (FC-121-1012). For three species, Illumina data were multiplexed across two PE150 lanes on an Illumina HiSeq 4000 and demultiplexed by Novogene (Sacramento, CA USA). For *F. olivaceus*, PE150 Illumina NovaSeq reads from one flow cell (2 lanes) were graciously provided by the Texas A&M Agrilife Research Sequencing Facility (College Station, TX USA).

Data Description

Whole genome sequencing data from individuals of four killifish species collected from ONT PromethION (Table 1) and Illumina (NovaSeq and HiSeq 4000) (Table 2) were deposited in the European Nucleotide Archive (ENA) under the study accession PRJEB29136. Deposited raw data are untrimmed and unfiltered. Reads corresponding to lambda phage were filtered from ONT PromethION data using the NanoLyse program from NanoPack (version 1.1.0; [44]). Porechop (version 0.2.3) was used to remove residual ONT adapters and NanoFilt (version 2.2.0; [44]) was used to filter reads with an average quality score >Q5. After filtering and adapter trimming, ONT data from the PromethION ranged from 30-45x coverage for each species. NanoPlot (version 1.10.0; [44]) was used for visualization of ONT read qualities.

Species	Bases called (Gb)	Cov. (x)	Average read length	Reads N50	Q>5 Bases called (Gb)	Q>5 Avg. read length	ONT signal Accession	ONT fastq Accession
F. xenicus	38.5	35	2,449	5,733; n = 1,373,426	36.42	2,699	ERR3385273	ERR3385269
F. nottii	33.4	30.4	6,480	12,995; n = 700,534	31.06	7,548	ERR3385275	ERR3385271
F. catenatus	40.3	36.6	1,699	3,439; n = 2,687,295	34.28	2,021	ERR3385274	ERR3385270
F. olivaceus	50.1	45.5	4,595	11,670; n = 987,921	45.97	5,365	ERR3385276	ERR3385272

Table 1. ONT data collected from each species. Coverage assumes the genome size of each species is 1.1 Gb, measured for *F. heteroclitus* (Reid et al. 2017). Untrimmed reads were deposited in the ENA under study PRJEB29136. Q>5 reads were used in subsequent assemblies.

Average quality scores for all Illumina data were consistently above Q30 (Figure 3A). Residual Nextera adapters and bases with low quality scores were removed from Illumina reads using Trimmomatic PE (version 0.38) with conservative parameters, which included removing bases from each read with a quality score below Q2 and required a minimum read length of 25 bases each [45].

For *F. xenicus* and *F. catenatus*, ONT read qualities ranged from Q5 (minimum cutoff) to Q14 with read lengths generally ranging from 10 bp to 100kb (Figure 3B,C). For *F. nottii and F. olivaceus*, ONT read qualities ranged from Q5 (minimum cutoff) to Q13 with read lengths ranging from 100 bp to 100kb (Figure 3D,E).

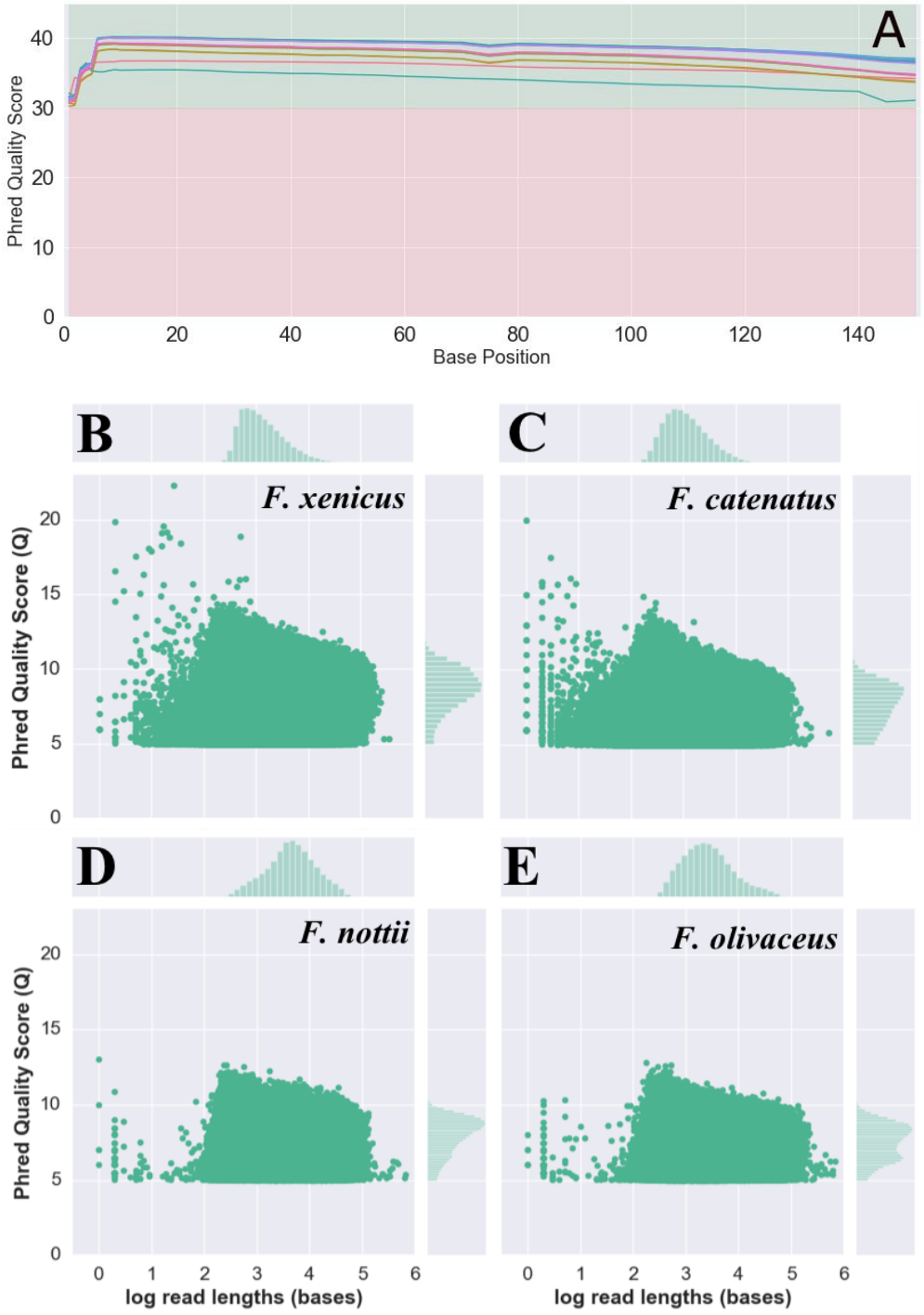

Figure 3. A) Quality score profiles for PE Illumina reads from *F. xenicus*, *F. catenatus*, *F. nottii* and *F. olivaceus.* For Illumina data, phred quality scores were consistently above Q30 across all reads. Average read quality scores (Q score) vs. read lengths for ONT PromethION from B) *F. xenicus*, C) *F. catenatus*, D) *F. nottii*, E) *F. olivaceus*.

Species	Platform	Reads (M)	Coverage (x)	FASTQ Accessions
F. xenicus	Illumina HiSeq	327.5	89.3	ERR3385278 ERR3385279
F. nottii	Illumina HiSeq	197	53.7	ERR3385282 ERR3385283
F. catenatus	Illumina HiSeq	316.5	86.3	ERR3385280 ERR3385281
F. olivaceus	Illumina NovaSeq	601.9	164	ERR3385284 ERR3385285

Table 2. Illumina data collected were all paired-end PE 150 reads. Coverage assumes 1.1 Gb genome size measured for *F. heteroclitus [40]*.

Draft Assemblies

As a comparison with assemblies using long ONT read data, Illumina data alone were assembled using ABySS version 2.1.5. While the BUSCO scores were consistently above 50% completeness [46], the number of contigs and contig N50 lengths of the Illumina-only assemblies were not acceptable for downstream use (Table 3).

Species	Bases in the Illumina-only assembly	N contigs	Avg length	Largest contig	N50	Illumina-only BUSCO C CS/CD/F/M
F. xenicus	1,283,257,056	5,195,861	246.98	71,596	2,571; n = 107,350	57.1% 56.4/0.7/33.3/9.6
F. catenatus	1,205,429,912	3,989,534	302.15	70,870	3,629; n = 80,839	53.8% 52.8/1.0/36.0/10.2
F. nottii	1,167,835,004	3,875,693	301.32	92,540	3,740; n = 72810	62.7% 61.7/1.0/27.4/9.9
F. olivaceus	1,252,948,998	4,509,089	277.87	70,765	3,670; n = 77136	65.7% 64.0/1.7/25.1/9.2

Table 3. Statistics for Illumina-only assemblies using ABySS (version 2.1.5) for each species. The BUSCO Eukaryota database (303 genes) was used to evaluate the completeness of each

assembly [46]. BUSCO numbers reported are percentage complete (C) followed by the percentages of complete single-copy (CS), complete duplicated (CD), fragmented (F), missing (M) out of 303 genes.

The ONT-only assemblies using the fuzzy de Bruijn graph assembler, wtdbg2 (version 2.3; [47]) had high contig N50 but low complete matches with the BUSCO Eukaryota database (Table 4). The assembler wtdbg2 took an average of 6.1 hours per assembly and required 59 GB RAM. The polishing tool pilon required an average of 65.99 hours and used 1.61 TB RAM. Following polishing with Illumina data using the pilon software tool version 1.23 [33], the BUSCO Eukaryota completeness scores increased to consistently greater than 90% (Table 4). A full table of BUSCO metrics can be found in Supplemental Table 1. Assemblies were deposited in the Open Science Framework (OSF) repository, https://doi.org/10.17605/osf.io/zjv86 and in zenodo record, https://doi.org/10.5281/zenodo.3251033.

Species	**Contigs**	**Contig N50**	**Assembly size (bases)**	**Complete BUSCO after wtdbg2 ONT-only** C CS/CD/F/M	**Complete BUSCO after pilon polishing with Illumina** C CS/CD/F/M
F. xenicus	5,621	888,041; n = 325	1,075,031,690	10.2% 10.2/0.0/11.6/78.2	90.5% 87.5/3.0/3.0/6.5
F. nottii	2,242	2,701,963; n = 95	1,081,276,623	28.4% 11.2/0.0/22.1/66.7	94.4% 92.1/2.3/1.0/4.6
F. catenatus	5,854	436,102; n = 780	1,163,592,740	11.2% 28.4/0.0/24.4/47.2	90.4% 88.4/2.0/2.6/7.0
F. olivaceus	2,622	2,669,230; n = 105	1,198,526,423	23.4% 23.4/0.0/25.7/50.9	92.1% 89.8/2.3/1.3/6.6

Table 4. ONT PromethION assemblies using the wtdbg2 version 2.3 assembler [47] followed by polishing with pilon version 1.23 [33]. Of interest is the dramatic improvement of the complete BUSCO metric after polishing with pilon. BUSCO numbers reported are percentage complete (C) followed by the percentages of complete single-copy (CS), complete duplicated (CD), fragmented (F), missing (M) out of the 303 genes in the BUSCO Eukaryota database [46].

Discussion

In this study, we collected 30-45x coverage of ONT data in combination with 50-160x coverage of Illumina PE150 sequencing data and generated draft genome assemblies for four species of *Fundulus* killifish. For these assemblies, the combination of ONT and Illumina data allowed us to generate highly contiguous assemblies with acceptable BUSCO results. The assemblies generated by ONT data alone were not acceptable for use because of the low BUSCO results, due to the high rate of ONT sequence errors. Polishing the ONT assemblies with the Illumina data did not improve contiguity of the assemblies, but served to correct errors, shown by the large boost in BUSCO scores relative to the ONT assemblies alone.

The qualities of the ONT data appeared to make a difference in the contig N50 metrics of the assemblies. *F. nottii* and *F. olivaceus* both had contig N50 >2Mb, while assemblies from *F. xenicus* and *F. catenatus* had contig N50 <1Mb. *F. xenicus* and *F. catenatus* had shorter average read lengths and reads N50, on average, compared to *F. nottii* and *F. olivaceus*. *F. nottii*, which had the lowest yield, had higher average read lengths and reads N50 compared to the other species. *F. olivaceus*, which had the highest yield, also had a high reads N50 and average read length. Therefore, when generating ONT data for draft genome assemblies, it might matter more to have a lower yield of longer reads than a higher yield of shorter reads.

While the ONT data collected were sufficient for genome assembly of these organisms, it is worth noting that our yields were lower than those advertised on the ONT website (~100 Gb from a single PromethION flow cell) (https://nanoporetech.com/products/promethion, accessed 06/12/2019), and read length N50 was shorter than expected based on DNA gel analysis. We observed that our samples were consistently not using pores as efficiently on the PromethION compared with other runs with samples isolated from human or mammalian samples. The reasons for this are not fully understood, but this could be due to brittle property of our hmw DNA.

The Vertebrate Genome Project (VGP 2018) lists standards for *de novo* long-range genome assembly that include PacBio long reads, 10x linked Illumina reads, Hi-C chromatin mapping and Bionano Genomics optical maps. These four types of data each have a high cost of generation as well as associated analysis time and computational costs. While chromatin capture and Hi-C methods produce chromosome-level assemblies of very high quality [48–51], this can significantly increase the cost of the genome sequencing project. Here, we report the pairing of high-quality short Illumina reads with error-prone long reads generated from the ONT PromethION platform to generate a draft assembly at a minimum cost. The qualities of these assemblies are not as high as compared to the standards recommended by VGP (2018) with the 3.4.2.QV40 phased metric, which requires the assembly to be haplotype phased with a minimum contig N50 of 1 million bp (1Mb), scaffold N50 of 10Mb, 90% of the genome assembled into

chromosomes and a base quality error of Q40, (VGP 2019). However, for many research purposes these assemblies are sufficient, considering the low cost and that we have a high-quality reference genome assembly for another species within the genus [40]. For *F. olivaceus* and *F. nottii*, draft assemblies using wtdbg2 [47] and pilon polishing with Illumina data [33] had contig N50 >1 Mb, which meets the minimum requirements for assemblies in downstream synteny analyses [11].

New software tools and methods for base-calling, assembling and analyzing noisy ONT long reads are being developed at a fast rate [52,53]. Because of this fast pace of software tool development for ONT data, standard operating procedures are not available. While we use these assemblies for their intended purpose of downstream comparative evolutionary analyses, the raw data are shared here with the intent that others may use them for tool development and as new workflow pipelines, algorithms, tools, and best practices emerge.

Conclusions

Sequencing data from the ONT PromethION and Illumina platforms combined can contribute to assemblies of eukaryotic vertebrate genomes (>1 Gb). These sequencing data from wild-caught individuals of *Fundulus* killifish species are available for use with tool development and workflow pipelines. Ongoing work from our group is comparing genomic content between *Fundulus* species to address questions about evolutionary mechanisms of osmoregulatory divergence.

Data re-use potential

We encourage use and re-use of these data for external analyses. This collection of whole genome sequencing data from the PromethION and Illumina platforms originates from wild-caught individuals of closely-related *Fundulus* killifish species, obtained for the purpose of downstream evolutionary genomic comparative analyses. These data add to the growing set of public data available from ONT PromethION sequencing platform [25,54] which can be used for developing base-calling and assembly algorithms with this type of data.

Availability of supporting data and materials

Raw data are available in the ENA under study PRJEB29136. Draft assembly data products and quality assessment reports are available in the OSF repository: https://doi.org/10.17605/osf.io/zjv86 and zenodo: https://doi.org/10.5281/zenodo.3251033. Scripts used for this analysis workflow are available at: https://github.com/johnsolk/ONT_Illumina_genome_assembly

List of abbreviations

BUSCO = Benchmarking Universal Single-Copy Orthologs
ENA = European Nucleotide Archive
hmw DNA = high molecular weight DNA
ONT = Oxford Nanopore Technologies
OSF = Open Science Framework
PE = paired end
VGP = Vertebrate Genome Project

Declarations

Ethical Approval
UC Davis IACUC protocol #17221

Consent for publication
Not applicable.

Competing Interests
The authors declare that they have no competing interests.

Funding
Gordon and Betty Moore Foundation to CTB under award number GBMF4551. IU-TACC Jetstream and PSC Bridges XSEDE allocations TG-BIO160028 and TG-MCB190015 to LKJ.

Author's Contributions

Sample extractions and library preparations were done by LKJ, RS, TG, JR. Project advising by CTB and AW. Manuscript writing and editing by LKJ, RS, TG, JR, LF, CTB, and AW.

Acknowledgements

We thank Dr. David Duvernell at Missouri University of Science & Technology and Dr. Jacob Schaefer at the University of Southern Mississippi for generously collecting and sending fish. A special thank you goes to Dr. Charlie Johnson and Dr. Richard Metz at Texas A&M University Agrilife Research Sequencing Facility for contributing Illumina NovaSeq data from *Fundulus olivaceus*. Thanks to the instructors and participants at PoreCamp USA (June 2017) for their helpful advice.

References

1. Mardis E, McPherson J, Martienssen R, Wilson RK, McCombie WR. What is finished, and why does it matter. Genome Res [Internet]. 2002;12:669–71. Available from: http://dx.doi.org/10.1101/gr.032102

2. Baker M. De novo genome assembly: what every biologist should know. Nat Methods [Internet]. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2012;9:333. Available from: https://doi.org/10.1038/nmeth.1935

3. Ekblom R, Wolf JBW. A field guide to whole-genome sequencing, assembly and annotation. Evol Appl [Internet]. 2014;7:1026–42. Available from: http://dx.doi.org/10.1111/eva.12178

4. Stemple DL. So, you want to sequence a genome. Genome Biol [Internet]. 2013;14:128. Available from: http://dx.doi.org/10.1186/gb-2013-14-7-128

5. Li F-W, Harkess A. A guide to sequence your favorite plant genomes. Appl Plant Sci [Internet]. 2018;6:e1030. Available from: http://dx.doi.org/10.1002/aps3.1030

6. Dominguez Del Angel V, Hjerde E, Sterck L, Capella-Gutierrez S, Notredame C, Vinnere Pettersson O, et al. Ten steps to get started in Genome Assembly and Annotation. F1000Res [Internet]. 2018;7. Available from: http://dx.doi.org/10.12688/f1000research.13598.1

7. Ip CLC, Loose M, Tyson JR, de Cesare M, Brown BL, Jain M, et al. MinION Analysis and Reference Consortium: Phase 1 data release and analysis. F1000Res [Internet]. 2015;4:1075. Available from: http://dx.doi.org/10.12688/f1000research.7201.1

8. Tyson JR, O'Neil NJ, Jain M, Olsen HE, Hieter P, Snutch TP. MinION-based long-read sequencing and assembly extends the Caenorhabditis elegans reference genome. Genome Res [Internet]. 2018;28:266–74. Available from: http://dx.doi.org/10.1101/gr.221184.117

9. Ebbert MTW, Jensen TD, Jansen-West K, Sens JP, Reddy JS, Ridge PG, et al. Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. Genome Biol [Internet]. 2019;20:97. Available from: http://dx.doi.org/10.1186/s13059-019-1707-2

10. Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, et al. Assessing the performance of the Oxford Nanopore Technologies MinION. Biomol Detect Quantif [Internet]. 2015;3:1–8. Available from: http://dx.doi.org/10.1016/j.bdq.2015.02.001

11. Liu D, Hunt M, Tsai IJ. Inferring synteny between genome assemblies: a systematic evaluation. BMC Bioinformatics [Internet]. 2018;19:26. Available from: http://dx.doi.org/10.1186/s12859-018-2026-4

12. Tyler AD, Mataseje L, Urfano CJ, Schmidt L, Antonation KS, Mulvey MR, et al. Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. Sci Rep [Internet]. 2018;8:10931. Available from: http://dx.doi.org/10.1038/s41598-018-29334-5

13. Pfeiffer F, Gröber C, Blank M, Händler K, Beyer M, Schultze JL, et al. Systematic evaluation of error rates and causes in short samples in next-generation sequencing. Sci Rep [Internet]. 2018;8:10950. Available from: http://dx.doi.org/10.1038/s41598-018-29325-6

14. Zeng Y, Martin CH. Oxford Nanopore sequencing in a research-based undergraduate course [Internet]. bioRxiv. 2017 [cited 2019 Jun 20]. p. 227439. Available from: https://www.biorxiv.org/content/10.1101/227439v1

15. Zaaijer S, Columbia University Ubiquitous Genomics 2015 class, Erlich Y. Using mobile sequencers in an academic classroom. Elife [Internet]. 2016;5. Available from: http://dx.doi.org/10.7554/eLife.14258

16. Ducluzeau A-L, Tyson JR, Collins RE, Snutch TP, Hassett BT. Genome Sequencing of Sub-Arctic Mesomycetozoean Sphaeroforma sirkka Strain B5, Performed with the Oxford Nanopore minION and Illumina HiSeq Systems. Microbiol Resour Announc [Internet]. 2018;7. Available from: http://dx.doi.org/10.1128/MRA.00848-18

17. Pomerantz A, Peñafiel N, Arteaga A, Bustamante L, Pichardo F, Coloma LA, et al. Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. Gigascience [Internet]. 2018;7. Available from: http://dx.doi.org/10.1093/gigascience/giy033

18. Boykin LM, Ghalab A, De Marchi BR, Savill A, Wainaina JM, Kinene T, et al. Real time portable genome sequencing for global food security [Internet]. bioRxiv. 2018 [cited 2019 Jun 20]. p. 314526. Available from: https://www.biorxiv.org/content/10.1101/314526v2

19. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature [Internet]. 2016;530:228–32. Available from: http://dx.doi.org/10.1038/nature16996

20. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc [Internet]. 2017;12:1261–76. Available from: http://dx.doi.org/10.1038/nprot.2017.066

21. Kafetzopoulou LE, Pullan ST, Lemey P, Suchard MA, Ehichioya DU, Pahlmann M, et al. Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. Science [Internet]. 2019;363:74–7. Available from: http://dx.doi.org/10.1126/science.aau9343

22. Schalamun M, Nagar R, Kainer D, Beavan E, Eccles D, Rathjen JP, et al. Harnessing the MinION: An example of how to establish long-read sequencing in a laboratory using challenging plant tissue from Eucalyptus pauciflora. Mol Ecol Resour [Internet]. 2019;19:77–89. Available from: http://dx.doi.org/10.1111/1755-0998.12938

23. Jain M, Koren S, Quick J, Rand AC, Sasani TA, Tyson JR, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads [Internet]. bioRxiv. 2017 [cited 2019 Jun 20]. p. 128835. Available from: https://www.biorxiv.org/content/10.1101/128835v1

24. Schmidt MH-W, Vogel A, Denton AK, Istace B, Wormit A, van de Geest H, et al. De Novo Assembly of a New Solanum pennellii Accession Using Nanopore Sequencing. Plant Cell [Internet]. 2017;29:2336–48. Available from: http://dx.doi.org/10.1105/tpc.17.00521

25. De Coster W, De Rijk P, De Roeck A, De Pooter T, D'Hert S, Strazisar M, et al. Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. Genome Res [Internet]. 2019; Available from: http://dx.doi.org/10.1101/gr.244939.118

26. Kim H-S, Jeon S, Kim C, Kim YK, Cho YS, Blazyte A, et al. Chromosome-scale assembly comparison of the Korean Reference Genome KOREF from PromethION and PacBio with Hi-C mapping information [Internet]. bioRxiv. 2019 [cited 2019 Jun 20]. p. 674804. Available from: https://www.biorxiv.org/content/10.1101/674804v1

27. Tan MH, Austin CM, Hammer MP, Lee YP, Croft LJ, Gan HM. Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (Amphiprion ocellaris) genome assembly. Gigascience [Internet]. 2018;7:1–6. Available from: http://dx.doi.org/10.1093/gigascience/gix137

28. Miller DE, Staber C, Zeitlinger J, Hawley RS. Highly Contiguous Genome Assemblies of 15 Drosophila Species Generated Using Nanopore Sequencing. G3 [Internet]. 2018;8:3131–41. Available from: http://dx.doi.org/10.1534/g3.118.200160

29. Cao MD, Nguyen SH, Ganesamoorthy D, Elliott AG, Cooper M, Coin LJM. Scaffolding and Completing Genome Assemblies in Real-time with Nanopore Sequencing [Internet]. bioRxiv. 2016 [cited 2019 Jun 20]. p. 054783. Available from: https://www.biorxiv.org/content/10.1101/054783v1

30. Giordano F, Aigrain L, Quail MA, Coupland P, Bonfield JK, Davies RM, et al. De novo yeast genome assemblies from MinION, PacBio and MiSeq platforms. Sci Rep [Internet]. 2017;7:3935. Available from: http://dx.doi.org/10.1038/s41598-017-03996-z

31. Austin CM, Tan MH, Harrisson KA, Lee YP, Croft LJ, Sunnucks P, et al. De novo genome assembly and annotation of Australia's largest freshwater fish, the Murray cod (Maccullochella peelii), from Illumina and Nanopore sequencing read. Gigascience [Internet]. 2017;6:1–6. Available from: http://dx.doi.org/10.1093/gigascience/gix063

32. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res [Internet]. 2017;27:737–46. Available from: http://dx.doi.org/10.1101/gr.214270.116

33. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One [Internet]. 2014;9:e112963. Available from: http://dx.doi.org/10.1371/journal.pone.0112963

34. Whitehead A. The evolutionary radiation of diverse osmotolerant physiologies in killifish (Fundulus sp.). Evolution [Internet]. 2010;64:2070–85. Available from: http://dx.doi.org/10.1111/j.1558-5646.2010.00957.x

35. Griffith RW. Environment and Salinity Tolerance in the Genus Fundulus. Copeia [Internet]. [American Society of Ichthyologists and Herpetologists (ASIH), Allen Press]; 1974;1974:319–31. Available from: http://www.jstor.org/stable/1442526

36. Durack PJ, Wijffels SE, Matear RJ. Ocean salinities reveal strong global water cycle intensification during 1950 to 2000. Science [Internet]. 2012;336:455–8. Available from: http://dx.doi.org/10.1126/science.1212222

37. Burnett KG, Bain LJ, Baldwin WS, Callard GV, Cohen S, Di Giulio RT, et al. Fundulus as the premier teleost model in environmental biology: opportunities for new insights using genomics. Comp Biochem Physiol Part D Genomics Proteomics [Internet]. 2007;2:257–86. Available from: http://dx.doi.org/10.1016/j.cbd.2007.09.001

38. Reid NM, Proestou DA, Clark BW, Warren WC, Colbourne JK, Shaw JR, et al. The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. Science [Internet]. 2016;354:1305–8. Available from: http://dx.doi.org/10.1126/science.aah4993

39. Oziolor EM, Reid NM, Yair S, Lee KM, Guberman VerPloeg S, Bruns PC, et al. Adaptive introgression enables evolutionary rescue from extreme environmental pollution. Science [Internet]. 2019;364:455–7. Available from: http://dx.doi.org/10.1126/science.aav4155

40. Reid NM, Jackson CE, Gilbert D, Minx P, Montague MJ, Hampton TH, et al. The landscape of extreme genomic variation in the highly adaptable Atlantic killifish. Genome Biol Evol [Internet]. 2017; Available from: http://dx.doi.org/10.1093/gbe/evx023

41. Ghedotti MJ, Davis MP. Phylogeny, Classification, and Evolution of Salinity Tolerance of the North American Topminnows and Killifishes, Family Fundulidae (Teleostei: Cyprinodontiformes). Fieldiana Life Earth Sci [Internet]. 2013;7:1–65. Available from: http://www.bioone.org/doi/abs/10.3158/2158-5520-12.7.1

42. Quick J. Ultra-long read sequencing protocol for RAD004 [Internet]. protocols.io; 2018 [cited 2019 Jun 20]. Available from: https://www.protocols.io/view/ultra-long-read-sequencing-protocol-for-rad004-mrxc57n

43. Wasko AP, Martins C, Oliveira C, Foresti F. Non-destructive genetic sampling in fish. An improved method for DNA extraction from fish fins and scales. Hereditas [Internet]. 2003;138:161–5. Available from: http://dx.doi.org/10.1034/j.1601-5223.2003.01503.x

44. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics [Internet]. 2018;34:2666–9. Available from: http://dx.doi.org/10.1093/bioinformatics/bty149

45. Macmanes MD. On the optimal trimming of high-throughput mRNA sequence data. Front Genet [Internet]. 2014;5:13. Available from: http://dx.doi.org/10.3389/fgene.2014.00013

46. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics [Internet]. 2015;31:3210–2. Available from: http://dx.doi.org/10.1093/bioinformatics/btv351

47. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2 [Internet]. bioRxiv. 2019 [cited 2019 Jun 20]. p. 530972. Available from: https://www.biorxiv.org/content/10.1101/530972v1

48. Olsen R-A, Bunikis I, Tiukova I, Holmberg K, Lötstedt B, Pettersson OV, et al. De novo assembly of Dekkera bruxellensis: a multi technology approach using short and long-read sequencing and optical mapping. Gigascience [Internet]. 2015;4:56. Available from: http://dx.doi.org/10.1186/s13742-015-0094-1

49. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nat Genet [Internet]. 2017;49:643–50. Available from: http://dx.doi.org/10.1038/ng.3802

50. Belser C, Istace B, Denis E, Dubarry M, Baurens F-C, Falentin C, et al. Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. Nat Plants [Internet]. 2018;4:879–87. Available from: http://dx.doi.org/10.1038/s41477-018-0289-4

51. Low WY, Tearle R, Bickhart DM, Rosen BD, Kingan SB, Swale T, et al. Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity. Nat Commun [Internet]. 2019;10:260. Available from: http://dx.doi.org/10.1038/s41467-018-08260-0

52. de Lannoy C, de Ridder D, Risse J. The long reads ahead: *de novo* genome assembly using the MinION. F1000Res [Internet]. 2017;6:1083. Available from: http://dx.doi.org/10.12688/f1000research.12012.2

53. Cali DS, Kim JS, Ghose S, Alkan C, Mutlu O. Nanopore Sequencing Technology and Tools for Genome Assembly: Computational Analysis of the Current State, Bottlenecks and Future Directions [Internet]. arXiv [q-bio.GN]. 2017. Available from: http://arxiv.org/abs/1711.08774

54. Nicholls SM, Quick JC, Tang S, Loman NJ. Ultra-deep, long-read nanopore sequencing of mock microbial community standards [Internet]. bioRxiv. 2018 [cited 2019 Jun 20]. p. 487033. Available from: https://www.biorxiv.org/content/10.1101/487033v2

55. De Bustos A, Cuadrado A, Jouve N. Sequencing of long stretches of repetitive DNA. Sci Rep [Internet]. 2016;6:36665. Available from: http://dx.doi.org/10.1038/srep36665

55. De Bustos A, Cuadrado A, Jouve N. Sequencing of long stretches of repetitive DNA. Sci Rep [Internet]. 2016;6:36665. Available from: http://dx.doi.org/10.1038/srep36665