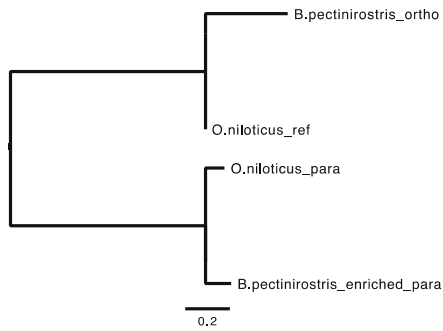


## Supplementary Material S1:

Table S1. Summary of sequencing statistics of 10 siniperoids.

Voucher number	Species name	Trimmed reads (bp) <sup>1</sup>
CL839_6	<i>S. kneri</i>	1,670,898
CL954	<i>S. kneri</i>	17,306,344
CL839_1	<i>S. kneri</i>	1,360,954
CL839_8	<i>S. kneri</i>	1,507,274
CL839_9	<i>S. kneri</i>	1,958,576
CL938_1	<i>S. chuatsi</i>	25,198,202
CL942_1	<i>S. chuatsi</i>	11,418,374
CL955_1	<i>S. chuatsi</i>	3,936,392
CL943_1	<i>S. chuatsi</i>	10,889,958
CL951_1	<i>S. chuatsi</i>	2,732,284

Danio\_rerio.8.33590631.33590784



Danio\_rerio.15.34647135.34646991

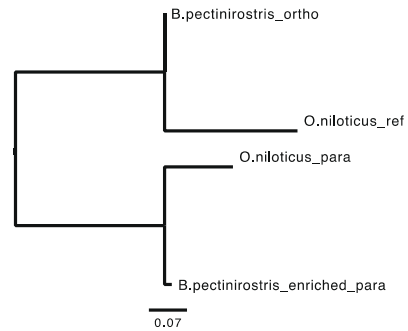


Fig. S1. ML trees of two genes showing the relationships among orthologous reference (*O. niloticus\_ref*) and misidentified paralogs generated by HybPiper in test 2 (*B. pectinirostris\_enriched\_para*). Orthologous sequence of *B. pectinirostris* (*B. pectinirostris\_ortho*) and paralogous sequence of *O. niloticus* (*O. niloticus\_para*) were retrieved from genome data.

The following material were also included online as separated files:

**Lepisosteus\_oculatus.fas**: reference sequences used in test 1

**Oreochromis\_niloticus.fas**: reference sequences used in test 2

**marker\_information.txt**: information of selected markers

**block\_pdis.pl**: scripts used to compute p-distance among flanks

**snp\_num.pl**: scripts used to count SNPs in coding and flanking regions

**test1\_result.zip**: assemblies of Assexon, CP, PHYLUCE and HybPiper in test 1

**test2\_result.zip**: assemblies of Assexon, CP, PHYLUCE and HybPiper in test 2

**flank\_results.tar.gz**: assemblies of 10 sinipercids

**consensus\_snp.all.filtered\_snps.vcf**: SNPs extracted from alignments of 10 sinipercids to consensus reference

## **Additional test on performance of exon assembly using four pipelines**

### **Methods**

Data of *Denticeps clupeioides* from Jiang *et al.*<sup>1</sup> were tested, which were enriched using baits designed on 4434 exon loci from genome of *Oreochromis niloticus*. Detailed information of captured data is listed in Table S2. This data was assembled with CP method in previous publication, so it could provide additional unbiased test for performance of the four pipelines. The same workflow and parameters described in the main text were used to assemble the reads. Then, recovered sequences were compared against the 4434 exon sequences of *O. niloticus* from Jiang *et al.*<sup>1</sup> and genome of *D. clupeioides* from VGP project (Genbank assembly accession: GCA\_900700375.1) to evaluate the completeness and accuracy of assemblies. SAR, SAG, CAR and CAG were calculated, and then these metrics were used to summarize the number of recovered loci, accurately assembled loci, perfectly assembled loci and paralogs generated for the four pipelines.

### **Results**

In this additional test, Assexon again recovered considerably more number of loci than PHYLUCE and HybPiper in all type of loci except paralogs. Assexon recovered consistently slightly more number of loci than CP in categories of recovered, accurately assembled and perfectly assembled loci (Table S3).

### **Reference**

1. Jiang J, Yuan H, Zheng X, et al. Gene markers for exon capture and phylogenomics in ray-finned fishes. *Ecol Evol.* Apr 2019;9(7):3973-3983.

Table S2. Summary statistics of gene-capture data on *Denticeps clupeioides*.

Species of sample	SRA accession number	Trimmed reads (bp) <sup>1</sup>	Species of reference targets (bp)	Number of target loci <sup>2</sup>	Divergence time (Myr) <sup>3</sup>	The closest species with genome available
<i>D. clupeioides</i>	SRR7903832	1,160,601,830	<i>O. niloticus</i>	4434	230	<i>D. clupeioides</i>

<sup>1</sup> The total base pairs of the reads after removing low quality bases

<sup>2</sup> Number of target loci in reference species

<sup>3</sup> Divergence time between the target species and reference

Table S3. Number of recovered, accurately assembled, perfectly assembled loci and paralogs produced using four pipelines on the data of *Denticeps clupeioides*.

	Recovered loci(%)	Accurately assembled loci (%)	Perfectly assembled loci (%)	Paralogs
Assexon	2105(47.5)	1831(41.3)	1434(32.3)	0
CP	1822(41.1)	1746(39.4)	1115(25.1)	0
PHYLUCCE	822(18.5)	808(18.2)	416(9.4)	3
HybPiper	753(17.0)	747(16.8)	615(13.9)	2