

a) Data acquisition (expanded dataset)

New sequence data: this study
202 taxa (156 pelagiarians, 46 outgroups)

<p>1,104 FishLife markers</p> <p>17 Legacy markers</p> <p>10 mtDNA markers</p>	<p>Exons & Exons + flanks</p>
---	--



Expanded data before QC:
this study + available data

<p>New sequence data</p> <p>156 pelagiarians, 46 outgroups</p> <hr style="width: 80%; margin: 5px auto;"/> <p>1,104 FishLife markers</p>	<p>Publicly available data</p> <p>104 pelagiarians, 38 outgroups</p> <hr style="width: 80%; margin: 5px auto;"/> <p>30-120 FishLife markers</p>
---	--

Dataset 1 - Exons: 344 taxa (260 pelagiarians, 84 outgroups) & 1,104 markers
Dataset 2 - Exons + flanks: 293 taxa (232 pelagiarians, 61 outgroups) & 1,104 markers



Expanded data after QC1:
this study + available data

<p>New sequence data</p> <p>155 pelagiarians, 46 outgroups</p> <hr style="width: 80%; margin: 5px auto;"/> <p>1,087 FishLife markers</p>	<p>Publicly available data</p> <p>74 pelagiarians, 38 outgroups</p> <hr style="width: 80%; margin: 5px auto;"/> <p>30-120 FishLife markers</p>
---	---

Dataset 3 - Exons: 313 taxa (229 pelagiarians, 84 outgroups) & 1,087 markers
Dataset 4 - Exons + flanks: 281 taxa (220 pelagiarians, 61 outgroups) & 1,087 markers



Expanded data after final QC:
this study + available data

<p>New sequence data</p> <p>140 pelagiarians, 43 outgroups</p> <hr style="width: 80%; margin: 5px auto;"/> <p>1,087 FishLife markers</p>	<p>Publicly available data</p> <p>20 pelagiarians, one outgroup</p> <hr style="width: 80%; margin: 5px auto;"/> <p>30-120 FishLife markers</p>
---	---

Dataset 5 - Exons: 204 taxa (160 pelagiarians, 44 outgroups) & 1,087 markers
Dataset 6 - Exons + flanks: 204 taxa (160 pelagiarians, 44 outgroups) & 1,087 markers

b) Subsets

New sequence data and UCE available data
44 taxa (42 pelagiarians, 2 outgroups)

<p>Exons</p> <p>964 FishLife markers</p> <p>422,964 sites</p>	<p>Exons + flanks</p> <p>377 FishLife markers</p> <p>422,946 sites</p>	<p>UCEs</p> <p>713 UCEs</p> <p>422,922 sites</p>
--	---	---

Correcting by number of sites for direct comparisons
alignment based analyses



New sequence data and UCE available data
44 taxa (42 pelagiarians, 2 outgroups)

<p>Exons</p> <p>377 FishLife markers</p>	<p>Exons + flanks</p> <p>377 FishLife markers</p>	<p>UCEs</p> <p>377 UCEs</p>
---	--	------------------------------------

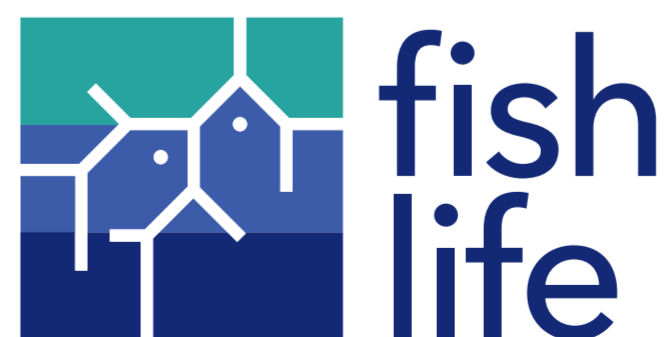
Correcting by number of genes for direct comparisons
of gene tree based analyses



Subsets data properties

44 taxa (42 pelagiarians, 2 outgroups)

<p>Exons</p> <p>422,964 sites</p> <p>Missing data: 9.6%</p> <p>PI sites: 102,797</p> <p>GC: 0.55</p>	<p>Exons + flanks</p> <p>422,946 sites</p> <p>Missing data: 50.3%</p> <p>PI sites: 119,593</p> <p>GC: 0.45</p>	<p>UCEs</p> <p>422,922 sites</p> <p>Missing data: 14.1%</p> <p>PI sites: 66,335</p> <p>GC: 0.43</p>
--	--	---



FishLife: genealogy and traits of living and fossil vertebrates that never left the water.
NSF initiative to resolve the Fish Tree of Life