

Additional file 3: Characterization of the high confidence novel genes

Table S13: Phylogenetic analysis of representative novel genes.

Species	EV ¹	ABHD14B	ATP5SL	B4GAT1	B9D2	CACNG7	COPZ1	ETFB	NAA10	OPA3	RPS19	RPS9	SLC50A1	SNRPD2	TXNIP	ZNRD1
Amino acid identity (%)																
<i>C. japonica</i> ²	40															
<i>A. platyrhynch</i> ³	84												51.3		91.7	
<i>H. leucocephalus</i> ⁴	102		66.1		81.4		98.0						57.1	100.0	91.5	57.1
<i>F. cherrug</i> ⁵	102	80.5				94.8	89.5						58.5		89.1	
<i>T. guttata</i> ⁶	102			67.0				71.5				99.5				
<i>C. livia</i> ⁷	102	78.6					96.3								89.9	
<i>C. mydas</i> ⁸	253	71.0	60.7				97.1						56.3		82.4	
<i>A. carolinensi</i> ⁹	278	68.6	45.0		71.6	84.3	95.5		82.6	82.5		98.9	55.3	100.0		47.1
<i>M. musculus</i> ¹⁰	321	64.3	42.1	56.9	69.9	89.5	96.6	66.0		74.8	95.9	99.4	50.3	100.0	78.5	49.6
<i>H. sapiens</i> ¹¹	321	64.5	45.7	56.2	69.9	89.5	96.6	66.0	82.5	75.7	95.9	99.4	48.0	100.0	78.6	48.7
<i>X. laevis</i> ¹²	356	60.4		60.0	65.3		91.5	64.1	81.9	76.3	91.0	97.8	52.2	98.3	74.7	
<i>D. rerio</i> ¹³	430	58.0		51.5	68.6		92.1	63.2	73.1	76.3	87.6	95.6	47.8	99.2	62.5	38.8
Correlation (aa identity vs. EV)		-1.0	-0.9	-0.9	-0.9	-0.7	-0.2	-1.0	-0.9	-0.5	-1.0	-0.7	-0.7	-0.5	-1.0	-0.9
GC-content (%) ¹⁴		70.5	66.9	69.5	70.6	66.9	63.2	69.1	60.0	68.1	65.3	68.3	70.4	62.0	67.0	61.5

¹Evolutionary distance from chicken in million years (MY), of the listed vertebrates was estimated using TimeTree web page: <http://www.timetree.org> [1]. ²*Coturnix japonica* (Japanese quail); ³*Anas platyrhynchos* (mallard); ⁴*Haliaeetus leucocephalus* (Bald eagle); ⁵*Falco cherrug* (Saker falcon); ⁶*Taeniopygia guttata* (Zebra finch); ⁷*Columba livia* (Rock dove); ⁸*Chelonia mydas* (Green sea-turtle); ⁹*Anolis carolinensis* (Green anole); ¹⁰*Mus musculus* (Mouse); ¹¹*Homo sapiens* (Human); ¹²*Xenopus laevis* (African clawed frog); ¹³*Danio rerio* (Zebrafish). ¹⁴GC-content in chicken.

One trinity transcript (*MRPL11*) was excluded from this study due to inclusion of an intron in *Chelonia mydas* [Accession no. XR_001463072], leading to a reduced similarity score.

1. Hedges SB, Marin J, Suleski M, Paymer M, Kumar S: **Tree of life reveals clock-like speciation and diversification.** Mol Biol Evol 2015, 32:835-845.

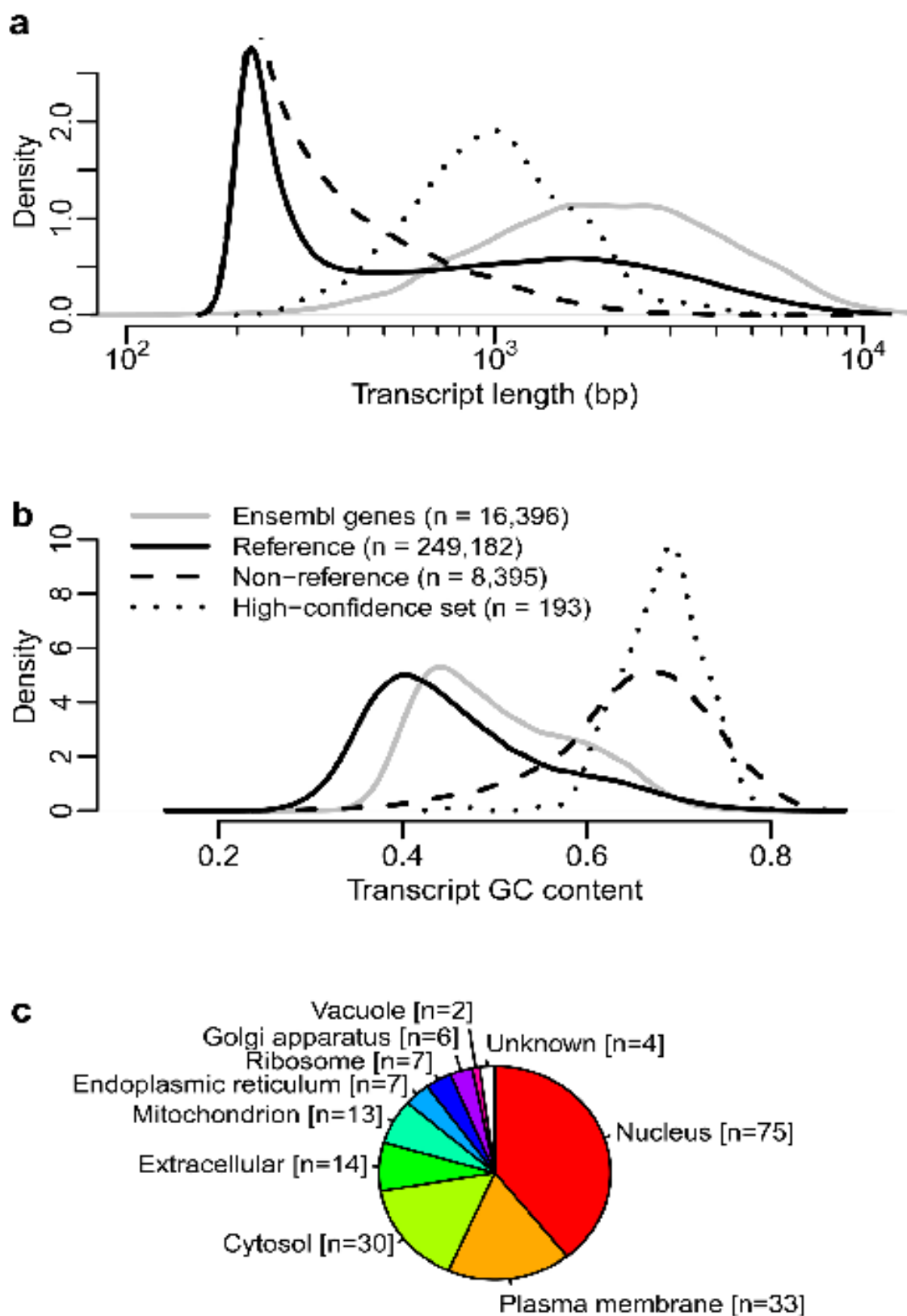


Figure S3: Characterization of the novel transcripts. **a** Length distribution of the Trinity

transcripts compared to protein-coding genes annotated in Ensembl. The ‘reference’ and ‘non-reference’ sets represent Trinity transcripts with or without alignment to the reference genome, respectively, and the ‘high-confidence set’ is the final set of novel genes selected from the non-reference set. **b** Distribution of GC-content in the Trinity transcripts compared to the protein-coding genes annotated in Ensembl as described in a. **c** Pie chart of the high confidence novel genes based on predicted subcellular localization. This prediction was made according to the mammalian orthologs using PANTHER, MGI and GOrilla software. The specific genes in each category are listed in Additional file 1: Table S7.

1. Hedges SB, Marin J, Suleski M, Paymer M, Kumar S: **Tree of life reveals clock-like speciation and diversification.** *Mol Biol Evol* 2015, **32**:835-845.