

Supplementary Material

SUPPLEMENTARY TABLE S1. AVERAGE INTRON SIZES
IN TEN LARGEST SCAFFOLDS OF COMMON CARP
ASSEMBLY COMPARED TO CORRESPONDING SYntenic
REGIONS IN ZEBRAFISH ZV9

<i>Scaffold Carp</i>	<i>Carp Average (bp)</i>	<i>Zv9 Average (bp)</i>	<i>Ratio Zebrafish/Carp</i>
1.1	1052	1480	1.41
2.1	1436	2230	1.55
3.1	1359	2200	1.62
4.1	1315	2101	1.60
5.1	875	1773	2.03
6.1	1190	5668	4.76
7.1	2657	2960	1.11
8.1	1112	3068	2.76
9.1	2553	3551	1.39
10.1	897	2764	3.08

The average ratio is 2.13 over all scaffolds.

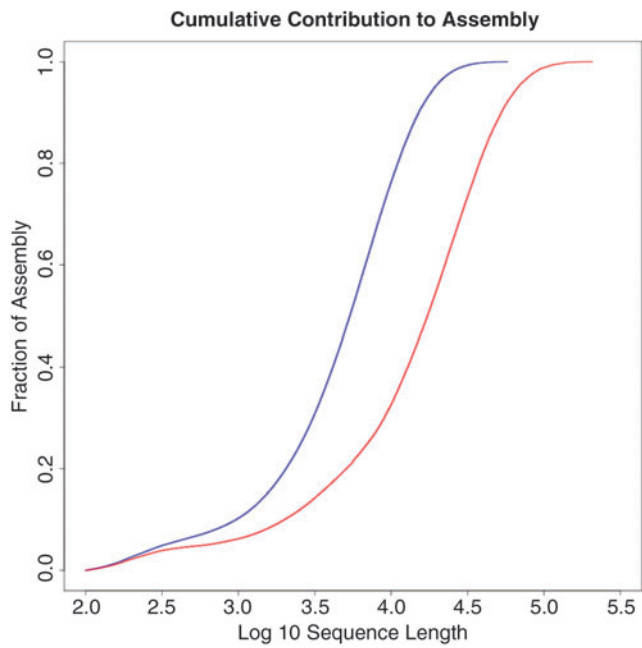
SUPPLEMENTARY TABLE S2. ZEBRAFISH GENES WITH ENSEMBL ANNOTATION FOR WHICH NO HOMOLOG
CAN BE IDENTIFIED IN THE CARP GENOME

<i>ENSEMBL Gene ID</i>	<i>EntrezGene ID</i>	<i>Unigene ID</i>	<i>Annotation</i>
ENSDARG00000058570	30268	Dr.76247	ccl1
ENSDARG00000075045	795785		cxcl-C1c
ENSDARG00000034222	553223	Dr.90211	HCLS1-binding protein 3 (LOC553223)
ENSDARG00000093249	561329		latrophilin 2 (sugar binding)
ENSDARG00000055679	100009640	Dr.80962	MTO1 (mitochondrial translation optimization 1 homolog)
ENSDARG00000058917	321023		nanor (nnr transcriptional regulator)
ENSDARG00000009890		Dr.75128	retrovirus ZFERV envelope protein (env)
ENSDARG00000088713	795554	Dr.75545	similar to bromodomain containing 7
ENSDARG00000068869	558896	Dr.114764	similar to Hspa12a
ENSDARG00000041433	562950		TPRXL (tetra-peptide repeat homeobox-like)
ENSDARG00000038271		Dr.84795	zgc:113293 (in situ: not spatially restricted)

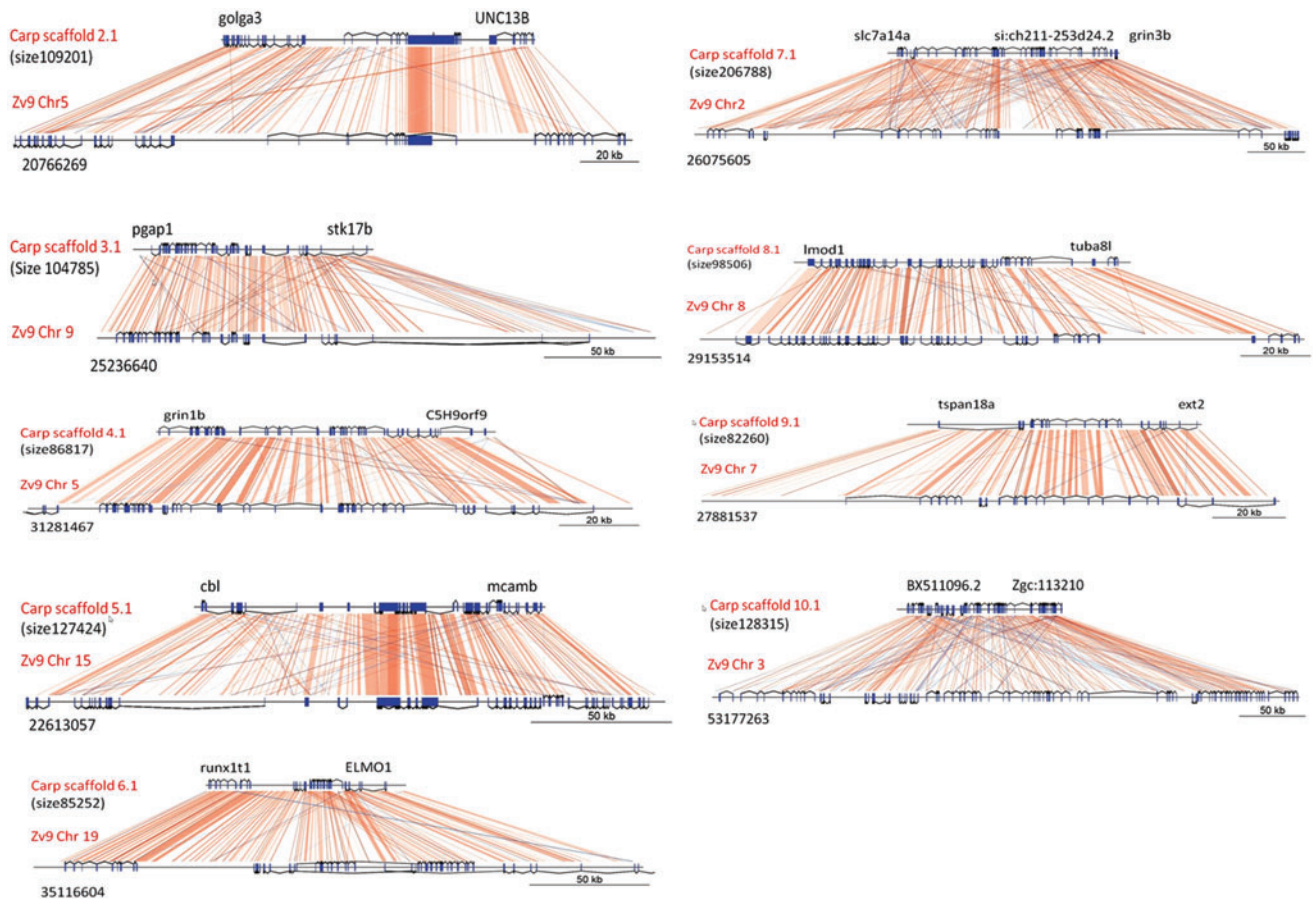
SUPPLEMENTARY TABLE S3. DESCRIPTION OF BEST SIGNIFICANT HITS (E-VALUE SMALLER THAN 0.01) RESULTING FROM TBLASTX ANALYSIS OF 47 CARP GENE PREDICTIONS FOR WHICH NO HOMOLOG CAN BE IDENTIFIED IN ZEBRAFISH ZV9 GENOME ASSEMBLY

Carp prediction query (bp)	Length (bp)	Best TBLASTX identifier	Best TBLASTX Length Target (bp)	Best tblastx hit species	BTH Query coverage	BTH E value	BTH score	General annotation from blast or do-main search	Notes from blast search
g58512	1179	XM_003454817	3047	<i>Oreochromis niloticus</i> (Nile tilapia)	44%	3e-39	61.9	rho guanine nucleotide exchange factor 26-like	Many rhogef hits, no ZF
g57951	1107	CP001102	Complete genome	<i>Candidatus Amoebophilus asiaticus</i> 5a2	43%	3e-26	73.9	ND	No eukaryote hit, no other good hits
g69595	1020	HQ021623	33396bp	<i>Sparus aurata</i> (gilthead seabream)	90%	2e-141	176	Similar to chaperonin, similar to TLR	TLR hit with <i>Oncorhynchus mykiss</i> DQ872852.1, no hit ZF
g51774	1005	CP001102	Complete genome	<i>Candidatus Amoebophilus asiaticus</i> 5a2	21%	9e-06	53.3	ND	No eukaryote hit, no other good hits
g7421	2871	X60391.1	1111	<i>Phaeococcus vulgaris</i>	9%	0.003	52.8	prolin-rich protein (PvPRP1)	Other proline rich hits also in animals, no hit ZF
g42526	942	FQ310506.2	Chromosome	<i>Dicentrarchus labrax</i> (European seabass)	93%	7e-134	180	Similar to variable lymphocyte receptor A, homology to homeobox	Many good hits with few function predictions, no hit ZF
g23091	900	XM_003447277	2778	<i>Oreochromis niloticus</i> (Nile tilapia)	65%	6e-36	74.8	C7orf63 homolog	Many homologs, no hit ZF
g68181	858	XM_003445937.1	2763	<i>Oreochromis niloticus</i> (Nile tilapia)	67%	3e-57	157	RING finger protein 17-like	Many homologs, no hit ZF
g52411	720	NM_001141647.1	2626	<i>Salmo salar</i>	58%	2e-44	187	Male-specific lethal 3-like 1	Many similar homologs, no hit ZF
g8791	687	XM_223231.5	3448	<i>Rattus norvegicus</i>	13%	6e-07	62.5	coiled-coil domain containing 158 (Cccl58)	Other coiled coil hits, no hit ZF
g13403	582	XM_003457514	2967	<i>Oreochromis niloticus</i> (Nile tilapia)	46%	3e-40	172	reversion-inducing-cysteine-richprotein with kazal motifs (RECK)	Many other hits with same predicted function, no hits ZF
g71092	1310	CP001102.1	Complete genome	<i>Candidatus Amoebophilus asiaticus</i> 5a2	23%	8e-15	65.0	ND	No other hits lower than 0.25 (pongo)
g69330	830	XM_001620354	975	<i>Nematostella vectensis</i> (starlet sea anemone)	45%	3e-35	137	Transmembrane, nonfunctional variable lymphocyte receptor (Eptatretus)	Many other homologs, indications for GPCR olfactory receptor, no hit ZF
g46173	738	XM_001624001	2832	<i>Nematostella vectensis</i> (starlet sea anemone)	69%	5e-53	197	Many homologs from a diversity of species, none with a predicted function	No hits ZF
g53183	648	XM_002612756.1	2054	<i>Branchiostoma floridae</i>	93%	7e-40	100	Many homologs from a diversity of species, all hypothetical	No hits ZF
g69331	1566	XM_001624001.1	2832	<i>Nematostella vectensis</i> (starlet sea anemone)	63%	7e-87	143	nonfunctional variable lymphocyte receptor A (VLRA)	Many other homologs, No hits ZF
g36022	1128	FQ212344.1	3022	<i>Rattus norvegicus</i>	12%	3e-5	57.9	APAIP-PTG-SIPD here described novel motive found in many mammals and chicken, C14orf37-like	No hits ZF

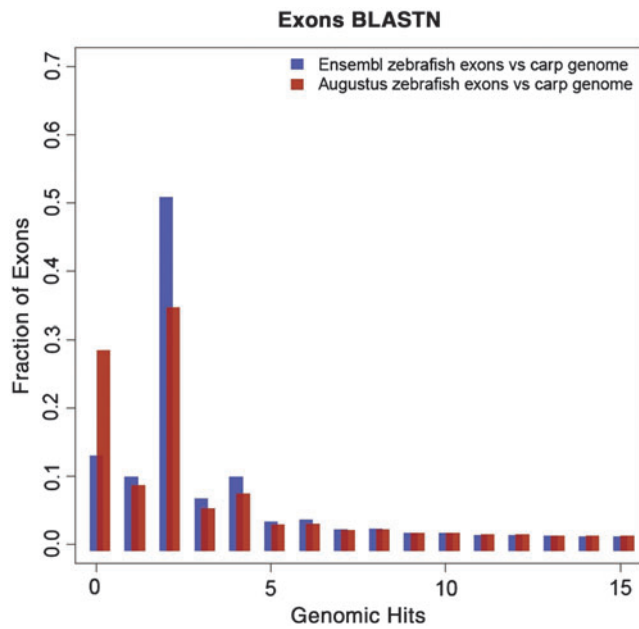
Manual annotation is based on evaluation of all blast hits and protein domain predictions. Abbreviations: ND, not determined; ZF, zebrafish.



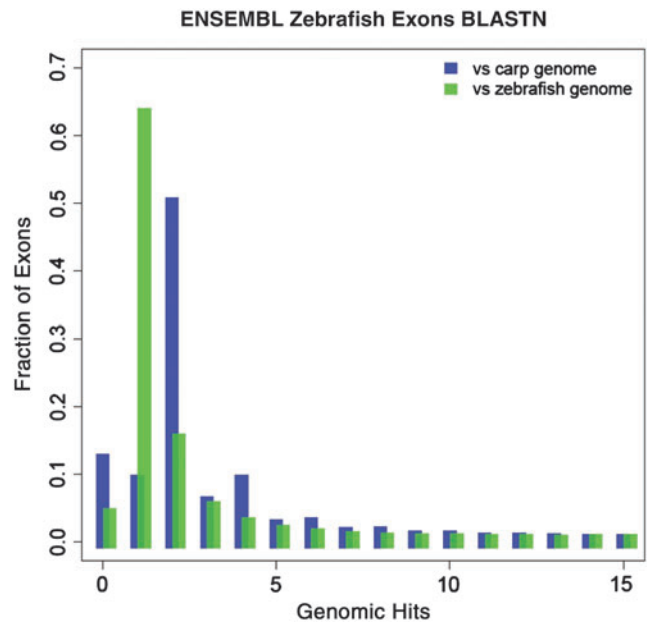
SUPPLEMENTARY FIG. S1. Contig/scaffold size distribution of the common carp genome assembly. The cumulative contribution (*y-axis*) of assembled fragments of a certain size (*x-axis*) to the total assembled genome size is shown in *blue* for the initial contigs, and in *red* for the final scaffolds. The fragment size at which the cumulative fraction is 0.5 represents the N50 of the assembly.



SUPPLEMENTARY FIG. S2. Synteny of 9 other largest carp scaffolds with the zebrafish Zv9 orthologs. The number indicated in the *right bottom of the panels* is the nucleotide start position in the Zv9 ENSEMBL database. The names of the genes indicated at the *borders of the scaffolds* are taken from the Zv9 ENSDARG annotations. The indicated gene structures were predicted with the program Augustus.



SUPPLEMENTARY FIG. S3. Analysis of multiplicity of blast hits. Results obtained using zebrafish exons predicted by Augustus instead of annotated by ENSEMBL as shown in Figure 5. ENSEMBL annotations in *blue*, Augustus prediction in *red*. Both show on average two best hits on the carp genome.



SUPPLEMENTARY FIG. S4. Analysis of multiplicity of blast hits. Supplement to Figure 5. As control, a BLASTN of zebrafish exons *versus* the zebrafish genome was performed (*green bars*) showing on average a single hit.