# 7 SUPPLEMENTARY DOCUMENT

## 7.1 Comparison with mammalian genome reconstruction methods

As discussed in Introduction, previous researchers have focused on macrosynteny for reconstructing the pre-TGD genome structure (Postlethwait *et al*., 2000; Naruse *et al*., 2004; Jaillon *et al*., 2004; Woods *et al*., 2005; Kohn *et al*., 2006; Kasahara *et al*., 2007; Muffato, 2010; see Muffato and Roest Crollius, 2008 for review).

On the other hand, various approaches have been employed for studying ancestral mammalian genomes: e.g., ancestral karyotypes were inferred by chromosome painting (see Ferguson-Smith and Trifonov, 2007 for review); rearrangement history and ancestral gene order were discussed using distance-based methods (Bourque and Pevzner, 2002;Bourque *et al*., 2004;Bourque *et al*., 2004); contiguous ancestral regions were reconstructed using homology-based methods (Ma *et al*., 2006; Chauve and Tannier, 2008; Gavranović *et al*., 2011); and large regions of ancestral genome sequences were reconstructed at single-nucleotide resolution (Blanchette *et al*., 2004; Paten *et al*., 2008; Diallo *et al*., 2010) by taking advantage of a large number of sequenced mammalian genomes. In addition, homology-based methods were used in reconstruction of ancestral amniote gene order (Ouangraoua *et al*., 2009; Ouangraoua *et al*., 2011).

However, those gene-order-based methods have not been applied to the reconstruction of pre-TGD gene order (Muffato and Roest Crollius, 2008; Jaillon *et al*., 2009). There are two primary reasons for this limitation: (1) the problem of pre-WGD gene-order reconstruction involves massive duplications and deletions, and consequently those gene-order reconstruction methods specifically designed for non-WGD genomes are not applicable in a straightforward manner (see El-Mabrouk and Sankoff, 2012 for review), and (2) microsynteny conservation (i.e., conservation of gene order and gene proximity) is substantially weaker in teleost than in mammals (Sémon and Wolfe, 2007; Hufton *et al*., 2008; Ravi and Venkatesh, 2008), partially due to massive gene loss after the TGD, which impedes gene-order reconstruction methods that rely on strong conservation of microsynteny (see Introduction in Gavranović *et al*., 2011). In sum, teleost genomes seem to be in a particularly challenging situation for gene-order- or gene-adjacency-based methods, and this is why we needed to develop the macrosynteny model.

## 7.2 Notation

In order to increase the readability of the main text, major symbols are explained in words in Supplementary Table 2. Definition of these variables can be found in Sections 2.2, 2.3, and 2.4.

### 7.2.1 *Calculation of expectations*

Let $X$ be a random variable having pdf $q$ and $f$ be an integrable function with $\mathbb{E}[|f(X)|] < \infty$. Then, the expected value of $f(X)$ is calculated as follows:

$$\mathbb{E}[f(X)] = \int f(x)q(x)dx, \tag{15}$$

where integration is over all possible values of $X$. See (Durrett, 2013) or (Williams, 1991) for more detail.

**Table 2.** Symbols in the macrosynteny model and their meanings.

| | |
|---|---|
| $X_{s,g} = k$ | Gene $g$ in non-WGD segment $s$ is assigned to pre-WGD chromosome $k$. |
| $Y_{s,g}^{t,d} = c$ | The $d$-th ortholog in post-WGD species $t$ (of gene $g$ in non-WGD segment $s$) is located on chromosome $c$. |
| $U_{s,k} = p$ | Genes in non-WGD segment $s$ are assigned to pre-WGD chromosome $k$ with probability $p$. |
| $V_{t,k,c} = p$ | A gene in post-WGD species $t$, which is orthologous to a non-WGD gene assigned to pre-WGD chromosome $k$, is located on chromosome $c$ with probability $p$. |
| $n_{t,c}^{s,g} = j$ | Gene $g$ in non-WGD segment $s$ has $j$ orthologs on chromosome $c$ in post-WGD species $t$. |

## 7.3 Derivation of the VBEM update formulas

### 7.3.1 *Variational M-step:*

First, we fix both $q_{\widehat{X}}$ and $q_{\widehat{V}}$ and then derive $q_{\widehat{U}}^* = \text{argmax}_{q_{\widehat{U}}} F(q_{\widehat{X},\widehat{\Theta}})$, the optimal $q_{\widehat{U}}$ that maximizes the negative free energy. For this purpose, we define a key quantity:

$$I(u) = \mathbb{E}[\log(p_U(u)p_{X|U}(\widehat{X}|u))]. \tag{16}$$

Note that $I(\widehat{U}) = \mathbb{E}[\log(p_U(\widehat{U})p_{X|U}(\widehat{X}|\widehat{U}))|\widehat{U}]$. (If this seems confusing, see (Durrett, 2013, Example 5.1.5) or (Williams, 1991, Section 9.10).) Then, substituting $p_{\Theta,X,Y}$ as Equation (7) and $q_{\widehat{\Theta},\widehat{X}}(\widehat{\Theta},\widehat{X}) = q_{\widehat{U}}(\widehat{U})q_{\widehat{V}}(\widehat{V})q_{\widehat{X}}(\widehat{X})$ and extracting terms that are not dependent on $\widehat{U}$ as $c_1$, we can transform the negative free energy given by Equation (9) as

$$F(q_{\widehat{X},\widehat{\Theta}}) = \mathbb{E}\left[I(\widehat{U}) - \log(q_{\widehat{U}}(\widehat{U}))\right] + c_1 \tag{17}$$

$$= -\text{KL}\left(q_{\widehat{U}}||J\right) + c_2, \tag{18}$$

where we defined function $J(u) = \exp(I(u))/\int\exp(I(u'))du'$ and constant $c_2 = \log(\int\exp(I(u'))du') + c_1$. It follows from Equation (18) that the negative free energy is maximized at $q_{\widehat{U}} = J$, which minimizes the KL divergence. Thus, by the definition of the model described in Sections 2.2 and 2.3 and writing the normalizing constant as $c_3 = -\log(\int\exp(I(u'))du')$, we have

$$\begin{aligned}
\log(q_{\widehat{U}}^*(u)) &= I(u) + c_3\\
&= \textstyle\sum_s \log(p_{U_s}(u_s)) + \sum_{s,g}\mathbb{E}\left[\log\left(p_{X_{s,g}|U_s}(\widehat{X}_{s,g}|u_s)\right)\right] + c_3\\
&= \textstyle\sum_s\left\{\log\left(\dfrac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)}\right) + \sum_k(\widehat{\alpha}_k^{(s)} - 1)\log(u_{s,k})\right\} + c_3,
\end{aligned} \tag{19}$$

where we defined variational parameters $\widehat{\alpha}_k^{(s)}$ by

$$\widehat{\alpha}_k^{(s)} = \alpha_k + \textstyle\sum_{g=1}^{G_s} q_{\widehat{X}_{s,g}}(k). \tag{20}$$

Since the right-hand side of Equation (19) is a sum of the logarithms of Dirichlet densities (cf. Equation (1)), we see that $q_{\widehat{U}}^*$ is factorized as $q_{\widehat{U}}^*(u) = \prod_{s=1}^{S} q_{\widehat{U}_s}^*(u_s)$, where $q_{\widehat{U}_s}^*$ is the Dirichlet density with parameters $\widehat{\alpha}^{(s)}$.

Second, with a similar argument for $q_{\widehat{U}}^*$, we derive $q_{\widehat{V}}^*$, the optimal $q_{\widehat{V}}$ that maximizes $F(q_{\widehat{X},\widehat{\Theta}})$. We have

$$\begin{aligned}
\log(q_{\widehat{V}}^*(v)) &= \mathbb{E}\left[\log\left(p_V(v)p_{Y|X,V}(y|\widehat{X},v)\right)\right] + c_4\\
&= \sum_{k,t}\left\{\log\left(\frac{\Gamma(\sum_c \beta_c^{(t)})}{\prod_c \Gamma(\beta_c^{(t)})}\right) + \sum_c(\widehat{\beta}_c^{(k,t)} - 1)\log(v_{k,t,c})\right\} + c_4,
\end{aligned} \tag{21}$$

where $c_4$ is a normalizing constant and we defined variational parameters $\widehat{\beta}_c^{(k,t)}$ by

$$\widehat{\beta}_c^{(k,t)} = \beta_c^{(t)} + \sum_{s=1}^{S} \sum_{g=1}^{G_s} q_{\widehat{X}_{s,g}}(k) n_{t,c}^{s,g}. \tag{22}$$

This indicates that $q_{\widehat{V}}^*$ is factorized as $q_{\widehat{V}}^*(v) = \prod_{k=1}^{K} \prod_{t=1}^{T} q_{\widehat{V}_{k,t}}^*(v_{k,t})$, where $q_{\widehat{V}_{k,t}}^*$ is the Dirichlet density with parameters $\widehat{\beta}^{(k,t)}$.

*7.3.2 Variational E-step:* Similar to the M-step, we fix $q_{\widehat{U}}$ and $q_{\widehat{V}}$ and derive $q_{\widehat{X}}^*$, the optimal $q_{\widehat{X}}$ that maximize $F(q_{\widehat{\Theta},\widehat{X}})$. Extracting terms that are not dependent on $\widehat{X}$ as $c_5$, and factorizing $p_{X|U}(x|u)$ and $p_{Y|X,V}(y|x,v)$ with respect to $s = 1, \ldots, S$, $g = 1, \ldots, G_s$, $t = 1, \ldots, T$, and $d = 1, \ldots, D_{s,g}^{(t)}$, we have

$$\begin{aligned}\log(q_{\widehat{X}}^*(x)) &= \mathbb{E}\left[\log\left(p_{X|U}(x|\widehat{U}) p_{Y|X,V}(y|x,\widehat{V})\right)\right] + c_5 \\ &= \sum_{s,g}\left(A_{s,g} + \sum_{t,d} B_{s,g}^{t,d}\right) + c_5, \end{aligned} \tag{23}$$

where we defined

$$A_{s,g} = \mathbb{E}\Big[\log\left(p_{X_{s,g}|U_s}(x_{s,g}|\widehat{U}_s)\right)\Big], \tag{24}$$

$$B_{s,g}^{t,d} = \mathbb{E}\Big[\log(p_{Y_{s,g}^{t,d}|X_{s,g},V_{X_{s,g},t}}(y_{s,g}^{t,d}|k,\widehat{V}_{k,t}))\Big]. \tag{25}$$

Then, assuming that $\widehat{U}_s$ follows the Dirichlet distribution with parameters $\widehat{\alpha}^{(s)}$, $A_{s,g}$ can be calculated as follows:

$$A_{s,g} = \mathbb{E}[\log(\widehat{U}_{s,x_{s,g}})] = \psi_0\left(\widehat{\alpha}_{x_{s,g}}^{(s)}\right) - \psi_0\left(\sum_{k=1}^{K} \widehat{\alpha}_k^{(s)}\right), \tag{26}$$

where $\psi_n(x) = \frac{d^{n+1}}{dx^{n+1}} \log(\Gamma(x)) = \frac{d^n}{dx^n} \frac{\Gamma'(x)}{\Gamma(x)}$ is the polygamma function. Similarly, writing as $k = x_{s,g}$ and $c = y_{s,g}^{t,d}$, we have

$$B_{s,g}^{t,d} = \mathbb{E}[\log(\widehat{V}_{k,t,c})] = \psi_0\left(\widehat{\beta}_c^{(k,t)}\right) - \psi_0\left(\sum_{i=1}^{C_t} \widehat{\beta}_i^{(k,t)}\right). \tag{27}$$

Taken together, $q_{\widehat{X}}^*(x)$ is factorized as $q_{\widehat{X}}^*(x) = \prod_s \prod_g q_{\widehat{X}_{s,g}}^*(x_{s,g})$, where $q_{\widehat{X}_{s,g}}^*(x_{s,g})$ can be calculated as follows with a normalizing constant $c_6$:

$$\log\left(q_{\widehat{X}_{s,g}}^*(x_{s,g})\right) = \psi_0\left(\widehat{\alpha}_{x_{s,g}}^{(s)}\right) + \sum_{t=1}^{T} \sum_{d=1}^{D_{s,g}^{(t)}} B_{s,g}^{t,d} + c_6. \tag{28}$$

## 7.4 Newton-Raphson method for hyper-parameter estimation

For each iteration, we estimate optimal hyper-parameters values that maximize the negative free energy as follows. First, focusing on the terms that involve with $\alpha$ and writing the other terms as $c_7$, we have

$$F(q_{\widehat{X},\widehat{\Theta}}) = \sum_{s=1}^{S} \mathbb{E}\left[\log\left(p_{U_s}(\widehat{U}_s)\right)\right] + c_7 \tag{29}$$

$$\begin{aligned} &= \sum_{s=1}^{S} \log\left(\Gamma(\textstyle\sum_{k=1}^{K} \alpha_k) / \prod_{k=1}^{K} \Gamma(\alpha_k)\right) \\ &\quad + \sum_{s=1}^{S} \sum_{k=1}^{K} (\alpha_k - 1) \mathbb{E}\left[\log(\widehat{U}_{s,k})\right] + c_7. \end{aligned} \tag{30}$$

Then, assuming that $\widehat{U}_s$ follows the Dirichlet distribution with parameters $\widehat{\alpha}^{(s)}$, we obtain the partial derivatives necessary for the Newton-Raphson updates as follows:

$$\begin{aligned}\frac{\partial F(q_{\widehat{X},\widehat{\Theta}})}{\partial \alpha_i} &= S\left\{\psi_0\left(\sum_{k=1}^{K} \alpha_k\right) - \psi_0(\alpha_i)\right\} \\ &\quad + \sum_{s=1}^{S}\left\{\psi_0\left(\widehat{\alpha}_i^{(s)}\right) - \psi_0\left(\sum_{k=1}^{K} \widehat{\alpha}_k^{(s)}\right)\right\}, \end{aligned} \tag{31}$$

$$\frac{\partial^2 F(q_{\widehat{X},\widehat{\Theta}})}{\partial \alpha_i \partial \alpha_j} = S\left\{\psi_1\left(\sum_{k=1}^{K} \alpha_k\right) - \delta_{i,j} \psi_1(\alpha_i)\right\}. \tag{32}$$

In the same way, partial derivatives with respect to beta parameters are derived as follows:

$$\begin{aligned}\frac{\partial F(q_{\widehat{X},\widehat{\Theta}})}{\partial \beta_i^{(t)}} &= K\left\{\psi_0\left(\sum_{c=1}^{C_t} \beta_c^{(t)}\right) - \psi_0\left(\beta_i^{(t)}\right)\right\} \\ &\quad + \sum_{k=1}^{K}\left\{\psi_0\left(\widehat{\beta}_i^{(k,t)}\right) - \psi_0\left(\sum_{c=1}^{C_t} \widehat{\beta}_c^{(k,t)}\right)\right\}, \end{aligned} \tag{33}$$

$$\frac{\partial^2 F(q_{\widehat{X},\widehat{\Theta}})}{\partial \beta_i^{(t)} \partial \beta_j^{(t)}} = K\left\{\psi_1\left(\sum_{c=1}^{C_t} \beta_c^{(t)}\right) - \delta_{i,j} \psi_1\left(\beta_i^{(t)}\right)\right\}. \tag{34}$$

From these equations we calculate the Newton-Raphson updates to obtain optimal hyper-parameter values (Press, 2007).

## 7.5 Initialization of the approximate distribution

The VBEM algorithm iteratively updates pdf $q_{\widehat{\Theta},\widehat{X}}$ to obtain refined approximation to the true posterior. The iteration starts from a pdf, which we initialize by using variance-minimizing clustering of non-WGD segments.

First, we exclude segments from the shortest one until the total number of excluded orthologs just exceeds $L$% of the total orthologs. We set $L = 10$ because short segments tend to become outliers and affect clustering results. Second, after removing short segments, individual segments are defined as distinct clusters. Third, we merge two clusters so that the sum of variance (defined below) over all clusters is minimized. We have chosen a variance-minimizing clustering algorithm because it is robust to outliers. This step is repeated until the number of clusters decreases to preassigned $K$. Forth, for a small $\epsilon > 0$ and $g = 1, \ldots, G_s$, we set $q_{\widehat{X}_{s,g}}(k) = 1 - (K-1)\epsilon$ if segment $s$ is a member of cluster $k$ and $q_{\widehat{X}_{s,g}}(k) = \epsilon$ otherwise. If segment $s$ was excluded in the first step, we set $q_{\widehat{X}_{s,g}}(k) = 1/K$ for all $g = 1, \ldots, G_s$. Fifth, $q_{\widehat{U}_s}$ and $q_{\widehat{V}_{k,t}}$ are defined to be the Dirichlet pdfs with parameters given by Equations (11) and (12).

The definition of variance is given as follows. For post-WGD species $t$, non-WGD segment $s$ is associated with a $C_t$-dimensional vector $n_t^{(s)} = (n_{t,1}^{(s)}, \ldots, n_{t,C_t}^{(s)})$, where $n_{t,c}^{(s)} = \sum_{g=1}^{G_s} n_{t,c}^{s,g}$ denotes the number of genes in chromosome $c$ of post-WGD species $t$ that are orthologous to genes in segment $s$. A cluster, denoted by $\mathcal{S}$, is a set of segments and is associated with $C_t$-dimensional vector $n_t^{(\mathcal{S})} = \sum_{s \in \mathcal{S}} n_{t,c}^{(s)}$. We define the distance between the center of cluster $\mathcal{S}$ and segment $s \in \mathcal{S}$ with respect to post-WGD species $t$ by their vector argument:

$$d_t(\mathcal{S}, s) = \arcsin \frac{n_t^{(\mathcal{S})} \cdot n_t^{(s)}}{|n_t^{(\mathcal{S})}| \cdot |n_t^{(s)}|}, \tag{35}$$

where $|\cdot|$ denotes the vector norm. Then we define the variance of cluster $\mathcal{S}$ as

$$d^{(\mathcal{S})} = \sum_{s \in \mathcal{S}} G_s \left\{\sum_{t=1}^{T} d_t(\mathcal{S}, s)/T\right\}^2. \tag{36}$$

### 7.6 Convergence criteria for $F(q_{\widehat{X},\widehat{\Theta}})$ and $\widehat{\alpha}_k^{(s)}$

The VBEM algorithm iteratively update the variational parameters until $F(q_{\widehat{X},\widehat{\Theta}})$ converges to a local maximum. For diagnosing convergence, we calculate $F(q_{\widehat{X},\widehat{\Theta}})$ as $F(q_{\widehat{X},\widehat{\Theta}}) = F_{\widehat{U}} + F_{\widehat{V}} + F_{\widehat{X}}$, where each term is given by

$$F_{\widehat{U}} = \mathbb{E}[\log(p_U(\widehat{U})p_{X|U}(\widehat{X}|\widehat{U})/q_{\widehat{U}}(\widehat{U}))] \tag{37}$$

$$= \sum_{s=1}^{S} \log\left(\frac{\Gamma(\sum_{k=1}^{K}\alpha_k)}{\prod_{k=1}^{K}\Gamma(\alpha_k)}\frac{\prod_{k=1}^{K}\Gamma(\widehat{\alpha}_k^{(s)})}{\Gamma(G_s+\sum_{k=1}^{K}\alpha_k)}\right), \tag{38}$$

$$F_{\widehat{V}} = \mathbb{E}[\log(p_V(\widehat{V})p_{Y|X,V}(y|\widehat{X},\widehat{V})/q_{\widehat{V}}(\widehat{V}))] \tag{39}$$

$$= \sum_{k=1}^{K}\sum_{t=1}^{T} \log\left(\frac{\Gamma(\sum_{c=1}^{C_t}\beta_c^{(t)})}{\prod_{c=1}^{C_t}\Gamma(\beta_c^{(t)})}\frac{\prod_{c=1}^{C_t}\Gamma(\widehat{\beta}_c^{(k,t)})}{\Gamma(\sum_{c=1}^{C_t}\widehat{\beta}_c^{(k,t)})}\right), \tag{40}$$

$$F_{\widehat{X}} = \mathbb{E}[\log(1/q_{\widehat{X}}(\widehat{X}))] \tag{41}$$

$$= -\sum_{s=1}^{S}\sum_{g=1}^{G_s}\sum_{k=1}^{K} q_{\widehat{X}_{s,g}}(k)\log(q_{\widehat{X}_{s,g}}(k)). \tag{42}$$

Then, $F(q_{\widehat{X},\widehat{\Theta}})$ is considered to be converged if the updated value of $F(q_{\widehat{X},\widehat{\Theta}})$, $R_1$, and the previous value, $R_2$, satisfy the following condition: $(R_1 - R_2)/|R_2| < 0.00001$.

Next, $\widehat{\alpha}_k^{(s)}$ is considered to be converged if the updated value of $\widehat{\alpha}_k^{(s)}$, $r_{k,1}$, and the previous value, $r_{k,2}$, satisfy the following condition: $\sum_k |r_{k,1} - r_{k,2}|/K < 0.00001$.

### 7.7 Reconstruction of pre-TGD gene order

We reconstructed pre-TGD gene order using ANGES (Jones *et al.*, 2012) and PMAG$^+$ (Hu *et al.*, 2014) in addition to the GapAdj analysis presented in Section 5. We obtained gene-order information from two sets of species: (1) human, medaka, and *Tetraodon*; and (2) human, mouse, dog, chicken, spotted gar, zebrafish, medaka, stickleback, and *Tetraodon*. We also made smaller-scale datasets from the two sets of species using only universal markers (i.e., genes present in all species).

In ANGES analysis, we compared all pairs between non- and post-TGD species for computing ancestral contiguous sets, and parameters were set as follows: markers_doubled=0, acs_sa=1, acs_sci=1, acs_mci=1, acs_weight=1, acs_correction=0, c1p_linear=1, cip_heuristic=1. The results were summarized below, which indicate that ANGES inferred a large number of short CARs having a small number of genes due to weak gene-order conservation between non- and post-TGD genomes.

**Table 3.** ANGES inferred a large number of short Pre-TGD CARs.

| Number of species | | Universal | Number | Number of genes | |
|---|---|---|---|---|---|
| non-TGD | post-TGD | markers | of CARs | Max | Median |
| 1 | 2 | yes | 874 | 17 | 2 |
| 1 | 2 | no | 1268 | 11 | 2 |
| 5 | 4 | yes | 492 | 32 | 3 |
| 5 | 4 | no | 2841 | 25 | 3 |

Next we used PMAG$^+$, which was available as a web server at http://www.geneorder.org. It returned reconstruction results only for the dataset with universal markers from the nine species. The reconstruction by PMAG$^+$ consisted of four large CARs having $957$, $767$, $893$, and $324$ genes, respectively. For assessing the quality of these CARs, we calculated the proportion of orthologs located on the two most syntenic medaka chromosomes as discussed in Section 5. The proportions for the four CARs were $0.177$, $0.181$, $0.199$, and $0.261$, respectively, suggesting that the individual CARs probably consist of falsely joined genes from many pre-TGD chromosomes.

Taken together, these reconstructions confirm the observation described in Section 5 that gene-order conservation between non- and post-TGD genomes is not sufficiently strong for reliable inference of large pre-TGD CARs.

## REFERENCES

Blanchette,M. *et al.* (2004) Reconstructing large regions of an ancestral mammalian genome in silico. *Genome Res.*, **14**, 2412–2423.

Bourque,G. and Pevzner,P.A. (2002) Genome-Scale Evolution: Reconstructing Gene Orders in the Ancestral Species. *Genome Res.*, **12**, 26–36.

Bourque,G. et al. (2004) Reconstructing the Genomic Architecture of Ancestral Mammals: Lessons From Human, Mouse, and Rat Genomes. *Genome Res.*, **14**, 507–516.

Bourque,G. et al. (2005) Comparative architectures of mammalian and chicken genomes reveal highly variable rates of genomic rearrangements across different lineages. *Genome Res.*, **15**, 98–110.

Chauve,C. and Tannier,E. (2008) A Methodological Framework for the Reconstruction of Contiguous Regions of Ancestral Genomes and Its Application to Mammalian Genomes. *PLoS Comput. Biol.*, **4**, e1000234.

Diallo,A.B. *et al.* (2010) Ancestors 1.0: a web server for ancestral sequence reconstruction. *Bioinformatics*, **26**, 130–131.

Durrett,R. (2010) Probability: Theory and Examples (Edition 4.1). Cambridge University Press, Cambridge; New York.

Ferguson-Smith,M.A. and Trifonov,V. (2007) Mammalian karyotype evolution. *Nat. Rev. Genet.*, **8**, 950–962.

Hu,F. *et al.* (2014) Probabilistic Reconstruction of Ancestral Gene Orders with Insertions and Deletions. *IEEE ACM T. Comput. Bi.*, **11**, 667–672.

Hufton,A.L. *et al.* (2008) Early vertebrate whole genome duplications were predated by a period of intense genome rearrangement. *Genome Res.*, **18**, 1582–1591.

Jaillon,O. *et al.* (2009) "Changing by doubling", the impact of Whole Genome Duplications in the evolution of eukaryotes. *C. R. Biol.*, **332**, 241?253.

Jones,B.R. *et al.* (2012) ANGES: reconstructing ANcestral GEnomeS maps. *Bioinformatics*, **28**, 2388–2390.

Kohn,M. *et al.* (2006) Reconstruction of a 450-My-old ancestral vertebrate protokaryotype. *Trends Genet.*, **22**, 203–210.

Ma,J. *et al.* (2006) Reconstructing contiguous regions of an ancestral genome. *Genome Res.*, **16**, 1557–1565.

Naruse,K. *et al.* (2004) A Medaka Gene Map: The Trace of Ancestral Vertebrate Proto-Chromosomes Revealed by Comparative Gene Mapping. *Genome Res.*, **14**, 820–828.

Paten,B. *et al.* (2008) Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res.*, **18**, 1829–1843.

Postlethwait,J.H. *et al.* (2000) Zebrafish Comparative Genomics and the Origins of Vertebrate Chromosomes. *Genome Res.*, **10**, 1890–1902.

Press,W.H. et al. (2007) Numerical Recipes: the art of scientific computing, Third Edition (C++). Cambridge University Press, Cambridge; New York.

Ravi,V. and Venkatesh,B. (2008) Rapidly evolving fish genomes and teleost diversity. *Curr. Opin. Genet. Dev.*, **18**, 544–550.

Sémon,M. and Wolfe,K.H. (2007) Rearrangement Rate following the Whole-Genome Duplication in Teleosts. *Mol. Biol. Evol.*, **24**, 860–867.

Williams,D. (1991) Probability with Martingales. Cambridge University Press, Cambridge; New York.

Woods,I.G. *et al.* (2005) The zebrafish gene map defines ancestral vertebrate chromosomes. *Genome Res.*, **15**, 1307–1314.

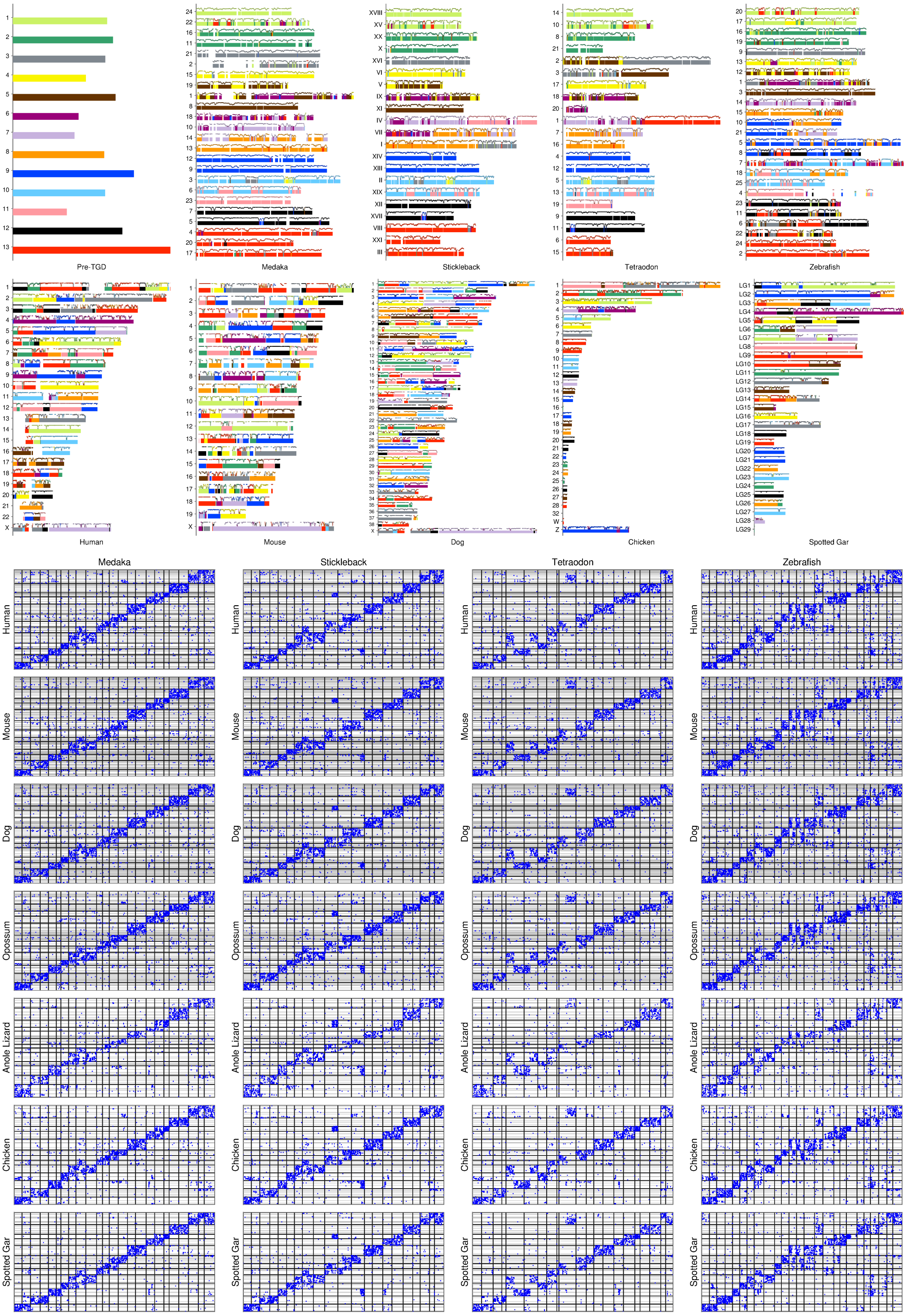

**Fig. 6.** Reconstruction with $K = 13$. The top half of this figure is identical to the figure presented in the main text. The bottom half shows ortholog distributions among post-TGD species (x-axis) and non-TGD species (y-axis). Post-TGD chromosomes were ordered as presented in the top: i.e., Ola24 to Ola17, Gac18 to Gac3, Tni14 to Tni15, and Dre20 to Dre2 (left to right). Non-TGD segments were assigned to pre-TGD chromosomes as described in the main text, and they were ordered along the x-axis from pre-TGD chromosomes 1 to 13 (bottom to top). Most clusters of blue dots distant from the diagonal lines are likely to indicate inter-chromosomal rearrangements in the post-TGD lineages.

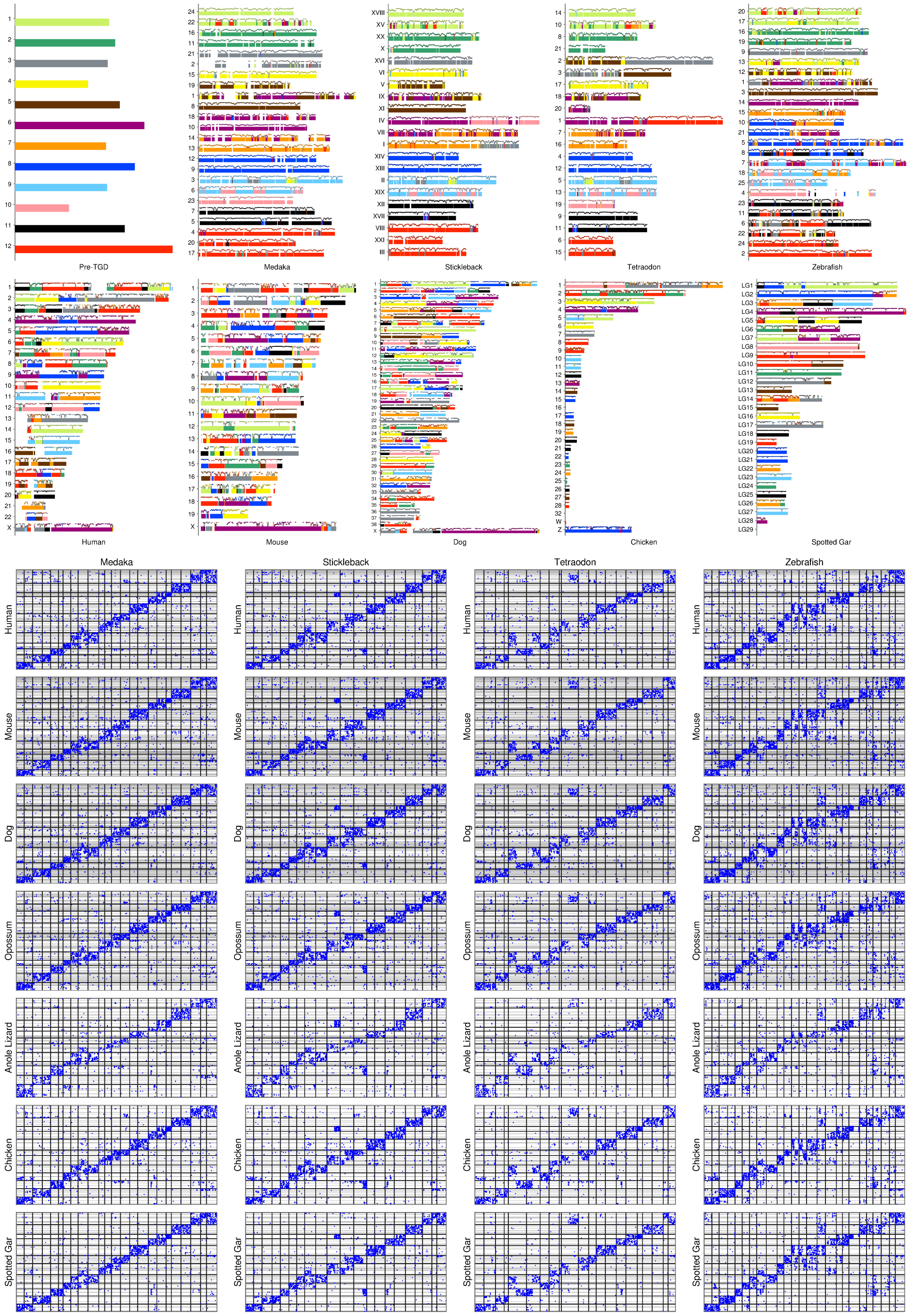

**Fig. 7.** Reconstruction with $K = 12$. The light purple pre-TGD chromosome in the $K = 13$ reconstruction was merged into the purple pre-TGD chromosome.

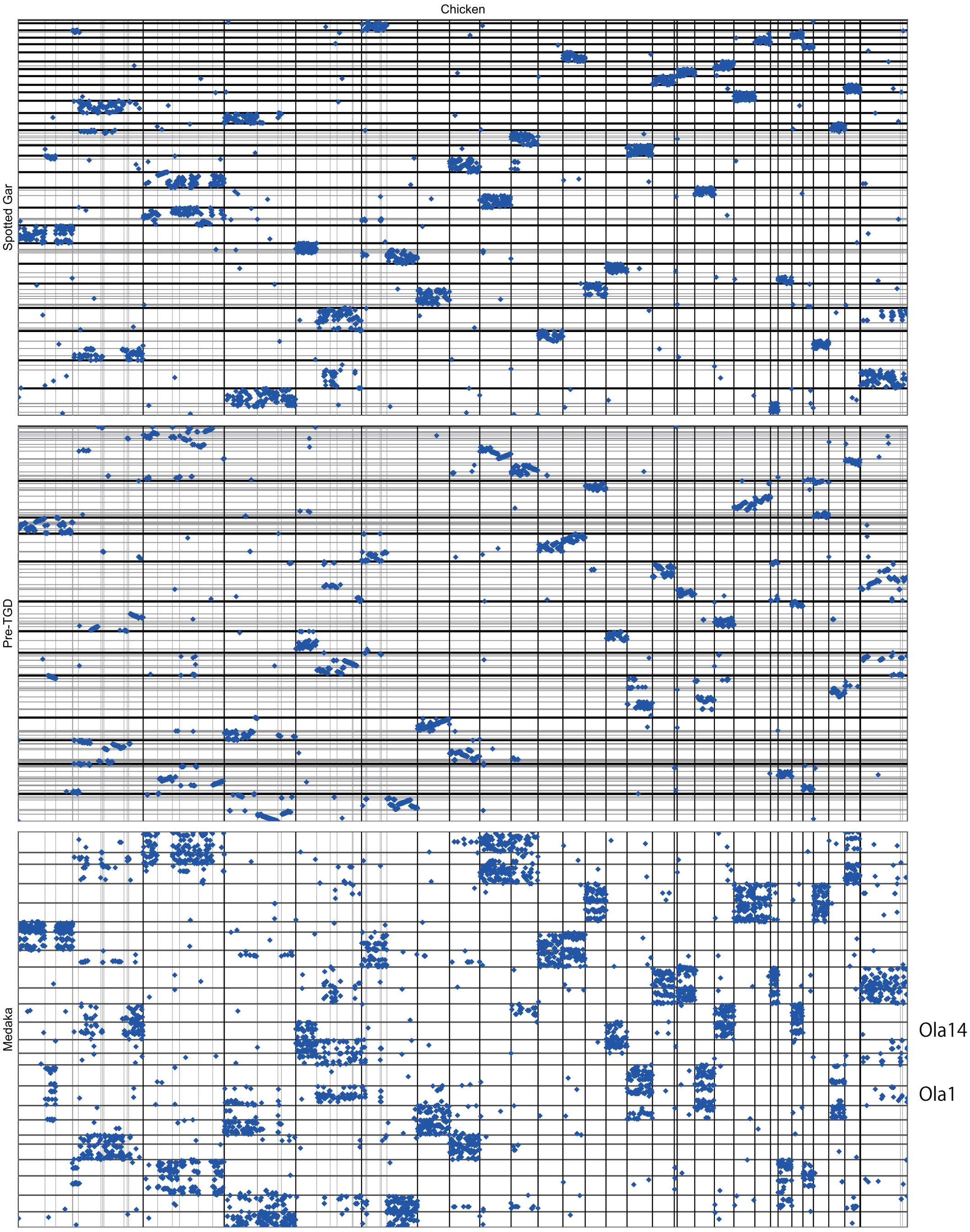

**Fig. 8.** Ortholog distribution among chicken, spotted gar, reconstructed pre-TGD ancestor ($K = 13$), and medaka. Black and gray lines indicate boundaries of chromosomes and conserved synteny blocks, respectively. The chicken chromosomes were ordered along the x-axis from Gga1 to GgaZ (left to right). The spotted gar and medaka chromosomes were ordered from Loc1 to Loc29 and from Ola24 to Ola17, respectively (bottom to top). The pre-TGD chromosomes, consisting of the 152 human segments, were ordered from chr1 to chr13 (bottom to top), and genes in the human segments were ordered as in the human genome. (The pre-TGD genome was represented by human segments and genes in this figure, considering the low coverage of the current version of chicken and spotted gar genomes.) The plot shows that (1) Ola1 and Ola14 have orthologs in non-overlapping regions in the chicken genome, and (2) many chicken microchromosomes retain one-to-one correspondence to the spotted gar chromosomes, but none of them were retained as single chromosomes in the pre-TGD genome.