

真骨鱼类 *DMRT* 基因家族的连锁结构及其系统发生

夏伟^{1,2} 周莉¹ 桂建芳¹

(1. 中国科学院水生生物研究所, 淡水生态与生物技术国家重点实验室, 武汉 430072; 2. 中国科学院研究生院, 北京 100039)

摘要: 利用已经报道的多个物种基因组序列资源, 我们通过比较真骨鱼类 (Teleost fish) 共有的 *DMRT* 1-5 基因所在染色体位置的基因组结构特征, 从基因组水平证明了真骨鱼类和四足动物的相应 *DMRT* 基因的直系同源关系, 揭示了 *DMRT* 4 和 *DMRT* 5 是因为真骨鱼类和四足动物的共同祖先发生了染色体重复事件而由同一基因分歧演变形成的。同时通过基因连锁分析, 探测到包括 *SNF* 2 和 *elavL* 家族基因在内的多个可能与 *DMRT* 基因功能密切相关的基因, 为进一步研究 *DMRT* 基因和它们之间可能存在的调控机制提供了新的线索。

关键词: 基因组; 连锁; 进化; 基因重复; 序列比较

中图分类号: Q343.2 文献标识码: A 文章编号: 1000-3207(2009)02-0254-10

基因家族的产生多数是由于基因重复, 而 DNA 重复是进化的主要动力之一。一个片段的拷贝数增加可由不同类型的重复所引起。长期以来学术界一直认为 DNA 重复是造成基因组大小变化的一个重要因素, 特别是全基因组重复或它的某一主要部分如一条染色体的重复可能会造成基因组大小突然急剧地增长。基因组重复事件曾在不同生物类群的进化中反复地被记录到, 而在植物^[1,2]、鱼类^[3-5]和两栖类^[6]中最为突出。基因重复的基本单元可以是完整的基因或一系列外显子, 甚至单个的外显子。当一个完整的基因重复产生两个活性完全相同的拷贝后, 两个拷贝便开始分别累积不同的突变。由于重复和变异从某一共同祖先基因分歧产生的一系列基因我们称之为一个基因家族 (Gene family)。

现有的研究表明, 脊椎动物具有的 DM 结构域 (*dsx* and *mb-3* related domain) 的相关基因构成了一个基因家族, 并且明显的聚为 7 类^[7], 它们所具有的唯一共同特征是均含有一个相似的 DM 结构域, 而不同类的包括同一个物种的 *DMRT* 基因在此区段以外的同源性则一般很低, 这就导致脊椎动物的 *DMRT* 基因命名比较混乱^[8], 我们也一直无法认定 *dsx*、*mb-3* 与脊椎动物的 *DMRT* 1 基因就是直系同源关系。虽然经过了近十年的研究, 也积累了不少

对 *DMRT* 基因的认识, 但是大多是关于其时空表达的研究, 对其作用分子机制的研究则较少, 相关的认识也很肤浅。新基因的产生方式和接下来的取舍而达成的平衡过程是我们了解一个基因家族所不能回避的问题。已有的关于 *DMRT* 家族进化机制的研究多是根据现存的基因种类来构建进化树, 然后通过进化树所反映的基因间的进化距离并结合相应物种的进化地位来推断基因家族的进化过程^[8], 但是这样往往会得出错误的结果^[9]。由于基因家族的演变必然反应在染色体组成的变化上面, 比如我们最近就在具天然性反转现象的石斑鱼中发现了一个可能是由于反转座事件而彻底丢失所有内含子的 *DMRT* 1 基因^[10], 同时考虑到现今世界上存在的动物被认为来自同一个祖先^[11], 因此本研究利用现有的多个物种的基因组数据库资源, 力图通过比对鱼类所具有的 5 种 *DMRT* 基因的基因组结构特征, 并尽可能参照四足动物以及相对低等的无脊椎动物基因组数据, 来试图更清晰地描绘鱼类 *DMRT* 家族的族谱。已有的关于 *DMRT* 基因连锁性的分析, 可能由于当时所能获得的基因组数据的限制都没有涉及到 *DMRT* 家族之外的其他基因^[7,12], 因此在本研究中我们进一步分析了与 *DMRT* 连锁的其他基因, 以此来预测可能在 *DMRT* 基因功能发挥中起重要作用的基因。

收稿日期: 2006-11-22 修订日期: 2007-10-18

基金项目: 国家 973 计划 (2004CB117401); 中国科学院知识创新工程重要方向项目 (KSCX2-YW-N-020) 资助

作者简介: 夏伟 (1979-), 男, 汉族, 安徽怀远人; 博士; 主要从事鱼类发育遗传学方面的研究。E-mail: xiewei7907@hotmail.com

通讯作者: 桂建芳, E-mail: jfg@ihb.ac.cn

1 材料与方法

1.1 基因组序列和潜在编码序列的获取 所有的基因组序列和潜在编码序列的获取都是通过 *ensembl* 网站 (<http://www.ensembl.org>) 的 *export data* 工具获取, *fasta* 格式蛋白序列的提取使用的是 *Emboss* 软件包的 *coderet* 程序^[13]。

1.2 序列的比较 全基因组的比较使用 *VISTA Browser*^[14] 直接调用已经比对好的结果 (<http://pipeline.lbl.gov/cgi-bin/gateway2>), 特定染色体片段的比较是使用 *wgVISTA* 的在线服务 (<http://genome.lbl.gov/cgi-bin/WGVistaInput>), 比较之前对重复序列的剔除是使用 *repeatmasker* 程序。蛋白序列的比对是使用本地 *blast* 程序包 (下载自 *NCBI* 主页) 的 *blastall* 程序。

1.3 连锁图以及染色体同源序列分布图的绘制 染色体连锁图是使用 *MAPMAKER3.0* (<http://linkage.rockefeller.edu/soft/mapmaker>), 染色体同源序列图是采用 *GenomePixelizer* (下载自 <http://www.atgc.org/GenomePixelizer/Download/GenoPixelizerDownload.html>), 所有编码氨基酸序列的相似性大于 40% 的基因都在图中标出。

2 结果与分析

2.1 *DMRT1-DMRT3-DMRT2* 的紧密连锁结构

在搜寻基因组数据库的过程中发现, 只要能同时找到这三个基因的脊椎动物物种, 包括人类、小鼠、非洲蛙、斑马鱼、青鳉以及红鳍东方鲀, *DMRT1-DMRT3-DMRT2* 都具有保守的同线性关系, 和已有的文献报道一致^[12], 暗示着辐鳍鱼类 (*Actinopterygii*) 与肉鳍鱼类 (*Sarcopterygii*) 分歧之前这三个基因就已经成簇出现了。因此我们以人类 9 号染色体短臂 (9p) 包含该同线性区域的大小为 3.6M 的片段为基准, 比较了人类和红鳍东方鲀、斑马鱼以及非洲蛙 *DMRT* 基因连锁群的 DNA 序列 (图 1)。该连锁群虽然只定位于人类 9 号染色体 831k 到 1.04M 的范围之内 (对应染色体 9p24.3), 但是这 4 个物种在对应 0.7—2.2M 的范围内都呈现高度的同源关系, 这就排除了这种结构在辐鳍鱼类和四足动物中独立形成的可能, 这 3 个 *DMRT* 基因必然在它们分歧之前就已经紧密连锁在一条染色体上了。但是这个区域两边保守的部分则分布在不同的染色体上 (图 1), 因此分歧之后不同物种的染色体必然经历了重排, 但这三个基因始终在一条染色体上相邻的位置, 暗

示着这种结构的重要性。我们进一步对附近区域人类与红鳍东方鲀可能编码的蛋白进行比对, 发现同样保守的基因还有 *SNF2* 基因 (这个位置的人类基因为 *hBRM*)、低密度脂蛋白相关受体 (*LDLR*) 以及 α 型锚蛋白重复单位结构域相关基因 (*Ankyrin repeat domain-containing gene 15* 简称 *ANKRD15*), 它们之间的线性关系如图 2 所示。有趣的是, 当我们把这三个基因在其他的模式生物基因组数据库中进行搜索时, 发现这种连锁情况同样存在于斑马鱼、非洲蛙、家鸡以及小鼠等物种中。

其中最引人注目的要算 *SNF2* 基因家族。我们更进一步在无脊椎动物玻璃海鞘 (*Ciona intestinalis*)、紫球海胆 (*Strongylocentrotus purpuratus*) 和果蝇 (*Drosophila melanogaster*) 的基因组中也发现了完整的 *SNF2* 基因的编码序列。由于海胆基因组的 scaffold 还没有定位到染色体上^[15], 我们现在还无法将其在染色体上所处的位置与 *DMRT* 基因进行比较。但在海鞘中, *SNF2* 基因定位在 5q 和 8q 染色体上, 并且这两条染色体都含有 *DMRT* 基因, 它们与 *SNF2* 基因在染色体上的距离也都不超过 1M。在果蝇中, *dsx* 基因位于 3R 染色体 3.75M 的位置, 就在邻近的 3.74M 的位置我们也发现了具有 *SNF2* 结构域的基因。进化地位如此悬殊的物种都含有与 *SNF2* 家族基因连锁 *DMRT* 基因的现象 (表 1) 似乎在向我们暗示它们之间存在着某种紧密不可分的关系。但是在线虫 *mab-3* 基因的附近我们没有找到 *SNF2* 基因, 不过鉴于在 *mab-3* 基因的表达调控中没有发现选择性剪接的现象, 这似乎又在暗示 *SNF2* 基因很有可能是在某些 *DMRT* 基因的选择性剪接过程中起重要的作用。

2.2 *DMRT4* 与 *DMRT5* 的进化

已有的进化树分析已经表明鱼类的 *DMRT4* 和 *DMRT5* 两个基因应该在鱼类和四足动物的祖先分歧之前就已经产生^[7]。考虑到鱼类特有的全基因组加倍事件^[16], 以及后来发生的各种染色体重排^[17, 18], 在染色体上与这两个基因邻近的序列或基因一般只有功能上紧密联系的才倾向于在进化距离如此远的不同物种间始终保持连锁关系, 因此我们对包含 *DMRT4* 和 *DMRT5* 两个基因的染色体片段进行了连锁分析。

由于我们不能从常用的模式生物如斑马鱼和青鳉的基因组中同时搜索到这两个基因, 而河豚 *DMRT4* 和 *DMRT5* 两个基因所在的 scaffold 却又都不超过 1M 大小, 因此我们引入这两个基因所在染

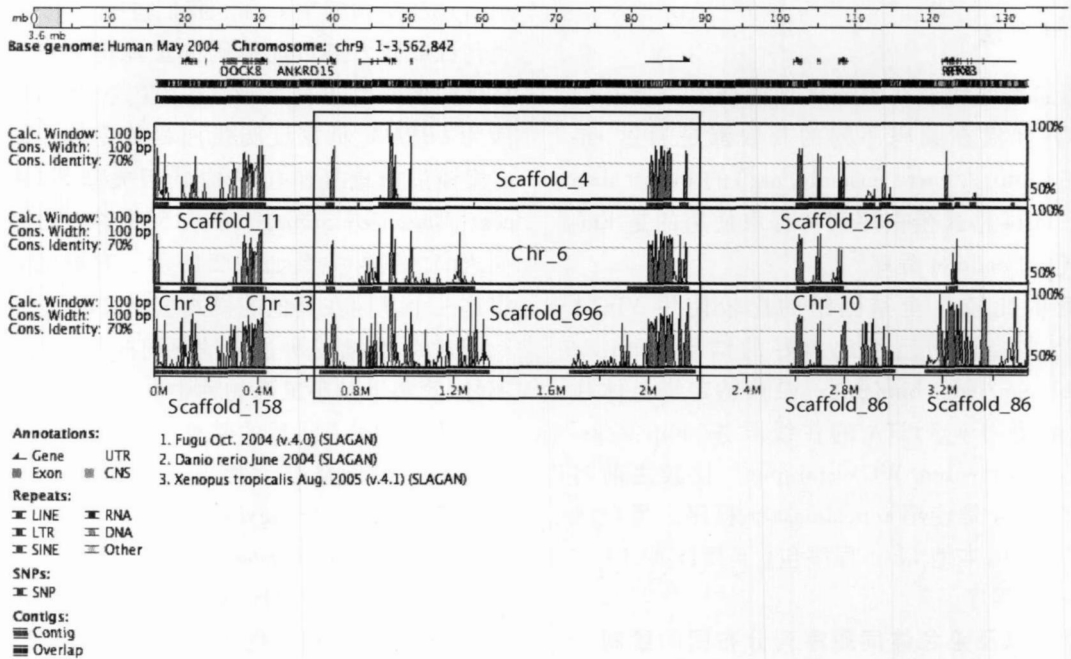


图 1 人类 *DMRT* 1-3-2 基因簇所在染色体片段与其他脊椎动物基因组的序列比较

Fig 1 The comparisons of human *DMRT* 1-3-2 cluster containing segments with other vertebrates' genomes

自上而下顺序分别为人类 9号染色体短臂端 3.6M 序列与红鳍东方鲀、斑马鱼和非洲蛙基因组同源序列的比较,大方框表示在此比较的 4 个物种间没有发生重排的人类 9号染色体区域,其内部标注的是该片的同源序列在其他 3 个物种基因组中的位置,同样标注的还有其他物种与该区域以外序列有同源性的片段在相应物种基因组中的位置。比较的方向分正反两个方向进行,横坐标表示对应人类序列的位置,纵坐标表示序列保守的水平,各种颜色代表的意义见图左下角注释

Comparing of the first 3.6M sequence of human chromosome 9p with the whole genome sequences of *Takifugu niphilus*, *Danio rerio* and *Xenopus tropicalis*. The names of chromosome and scaffold shown in the large box represent the segments that have not experienced rearrangements between these four species, and the names of other sequences showing similarities in this analysis are also indicated. The comparisons were performed on both strands of DNA. The horizontal axis shows the position of chromosome 9p sequence, and the level of conservation is displayed on the vertical axis. The meanings of different colors are shown in the annotation at the lower left corner in this figure.

染色体序列相对较为完整的三刺鱼 (*Gasterosteus aculeatus*) 来作为与其他物种进行比较分析的基础 (Base genome)。如图 3 所示,我们以三刺鱼 *DMRT* 4 和 *DMRT* 5 分别为中心,对其上下游各取 2.5M 大小的基因组序列作为基准,比较了三刺鱼与其他鱼类、两栖类以及人类基因组同源部分的相似程度,结果表明三刺鱼的这段序列与其他鱼类的相似性最高,而与人类相应区域的相似性最低,并且 *DMRT* 5 基因所在的周边序列较 *DMRT* 4 基因的要保守得多 (图 3A 1-3 和图 3B 1-3)。

比较人类和三刺鱼相应保守位置的潜在氨基酸编码序列发现,与 *DMRT* 4 基因连锁的基因只有 *elavL2* (embryonic lethal abnormal vision, *Drosophila* like 2), 它属于 *elavHu* 基因家族。同样有趣的是,当将 *elavL2* 的序列输入不同物种的基因组数据库进行搜索时发现,脊椎动物的 *DMRT* 4 和 *DMRT* 5 基因旁边都有 *elavL* 基因 (表 1), 并且 *DMRT* 4 和 *DMRT* 5

分别与 *elavL2* 和 *elavL4* 相连锁; 在无脊椎动物海鞘、果蝇和线虫的一些 *elavL* 基因旁边也可找到 *DMRT* 基因。譬如,海鞘 5q 染色体上与 *SVF2* 连锁的 *DMRT* 基因、果蝇的 *DMRT1E* 以及线虫的 *dm d-5* 等。

通过对非洲蛙及三刺鱼 *DMRT* 4 和 *DMRT* 5 基因在基因组中所处位置的潜在编码基因的进一步线性分析表明,三刺鱼 *DMRT* 4 和 *DMRT* 5 基因的附近存在大量同源基因 (图 4 黑线所示), 与图 3A-4 结果一致,它们的相似性甚至超过非洲蛙和三刺鱼的 Group VII 与 Scaffold_78 (被认为有着共同的起源) 之间的相似程度。与此类似的是非洲蛙的 Scaffold_78 和 Scaffold_1 之间的相似性更高 (图 5), 但是大多为零散的非编码序列,当我们将它们的潜在编码基因进行氨基酸比对时,仅仅发现了 5 个同源基因 (图 4)。从图 3B 和图 4 中可以清楚地观察到 Group VIII 与 Scaffold_1 之间的高度相似关系,而且 Group VII 与

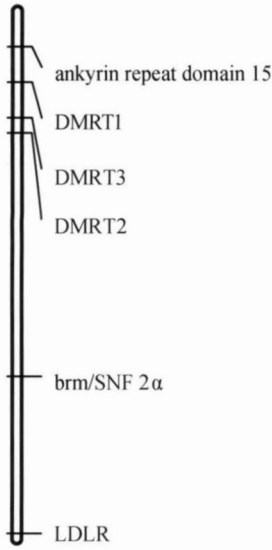


图 2 脊椎动物 *DMRT* 1-3-2 基因簇连锁基因线性关系的染色体模式图

Fig. 2 The chromosomal model shows the conserved synteny of *DMRT* 1-3-2 linked genes from vertebrates

右侧为基因名称, 之间距离表示各个基因在人类 9号染色体上的相对远近, 但没有严格的比例关系

Right of chromosome is the gene names and the distances shown here represent the relative positions of these genes on the real chromosome but without strict proportion

表 1 不同物种与 *DMRT* 基因连锁的 *SNF 2* 以及 *elavL* 基因在基因组中的位置

Tab. 1 The genomic positions of *SNF 2* and *elavL* genes in different species

| | <i>SNF 2</i> | <i>elavL</i> |
|--------------------------------|------------------------------|----------------------------|
| <i>Takifugu rubripes</i> | 1. 881- 1. 901 Sca_4 | |
| <i>Gasterosteus aculeatus</i> | 19. 1765 - 191768 group XIII | 5. 041- 5. 050 GroupVII |
| | | 13. 998- 14. 028 GroupVIII |
| <i>Danio rerio</i> | 41. 811- 41. 944 Chr5 | |
| <i>Gallus gallus</i> | 27. 123- 27. 225 ChrZ | |
| <i>Xenopus tropicalis</i> | 0. 378 - 0. 477 Sca_696 | 4. 584 - 4. 611 Sca_1 |
| | | 1. 280- 1. 326 Sca_78 |
| <i>Mus musculus</i> | 26. 687- 26. 843 Chr19 | 90. 745- 90. 800 chr4 |
| | | 109. 704- 109. 749 chr4 |
| <i>Homo sapiens</i> | 12. 005- 21. 836 Chr9 | 50. 383- 50. 439 chr1 |
| | | 23. 683- 23. 752 chr9 |
| <i>Drosophila melanogaster</i> | 3. 746 - 3. 750 Chr3R | 12. 777 - 12. 779 ChrX |
| <i>C. elegans</i> | | 9. 578 - 9. 580 ChrII |
| <i>Ciona Intestinalis</i> | 4. 980- 4. 989 5q | |
| | 4. 7083- 4. 7084 8q | 3. 346 - 3. 351 5q |

Scaffold_78的相似性明显高于与 Scaffold_1的相似性 (Group VII与 Scaffold_1的相似基因仅限于 *ELAV* 和 *DMRT* 基因), 上述结果表明两个物种的 *DMRT* 4和 *DMRT* 5基因分别为直系同源的关系。在同一物种中, *DMRT* 4基因和 *DMRT* 5基因所处染色体片段的高度相似性暗示它们可能来源于同一个祖先, 是部分或者全部的 DNA 重复之后的趋异进化产生目前两个不同的 *DMRT* 基因的, 而这个重复事件必然是出现在真骨鱼类和爬行类发生分歧以前的祖先身上。

从图 3和图 4中我们还可以看到, 相对于 *DMRT* 4同源基因来说, *DMRT* 5基因附近序列更加保守, 并且多集中在与 *DMRT* 5基因非常邻近的范围 (图 4红色线条所示, 非洲蛙 0.6M 和三刺鱼 0.25M 的范围)。在这段区域内, 非洲蛙和三刺鱼所含的基因数分别只有 11和 16个, 但是它们之间相似性超过 70% 的基因就达 10对之多, 占到整个 5M 范围相似基因数目的 56%, 并且同源比对的结果表明这 10对基因都是功能性基因。这表明 *DMRT* 5附近的基因与 *DMRT* 5同样具有高度保守性。这些基因之间究竟存在什么关系, 是否都与脑或性腺的发育与分化相关还有待深入的研究。

3 讨论

传统的观念认为, 性别决定与分化相关基因的进化速率非常快。作为第一个在无脊椎和脊椎动物间皆保守存在的性别决定和分化相关的基因,

*DMRT*基因家族一经发现便引起了广泛的关注。已有的研究表明, 在脊椎动物中发现的 *DMRT* 基因的数量并不多, 基本上可以聚为 7类。虽然鱼类历经比四足动物多一次的染色体加倍事件^[4, 5, 19], 但目前已报道了的鱼类共有的 *DMRT* 基因也只有 5个^[7]。

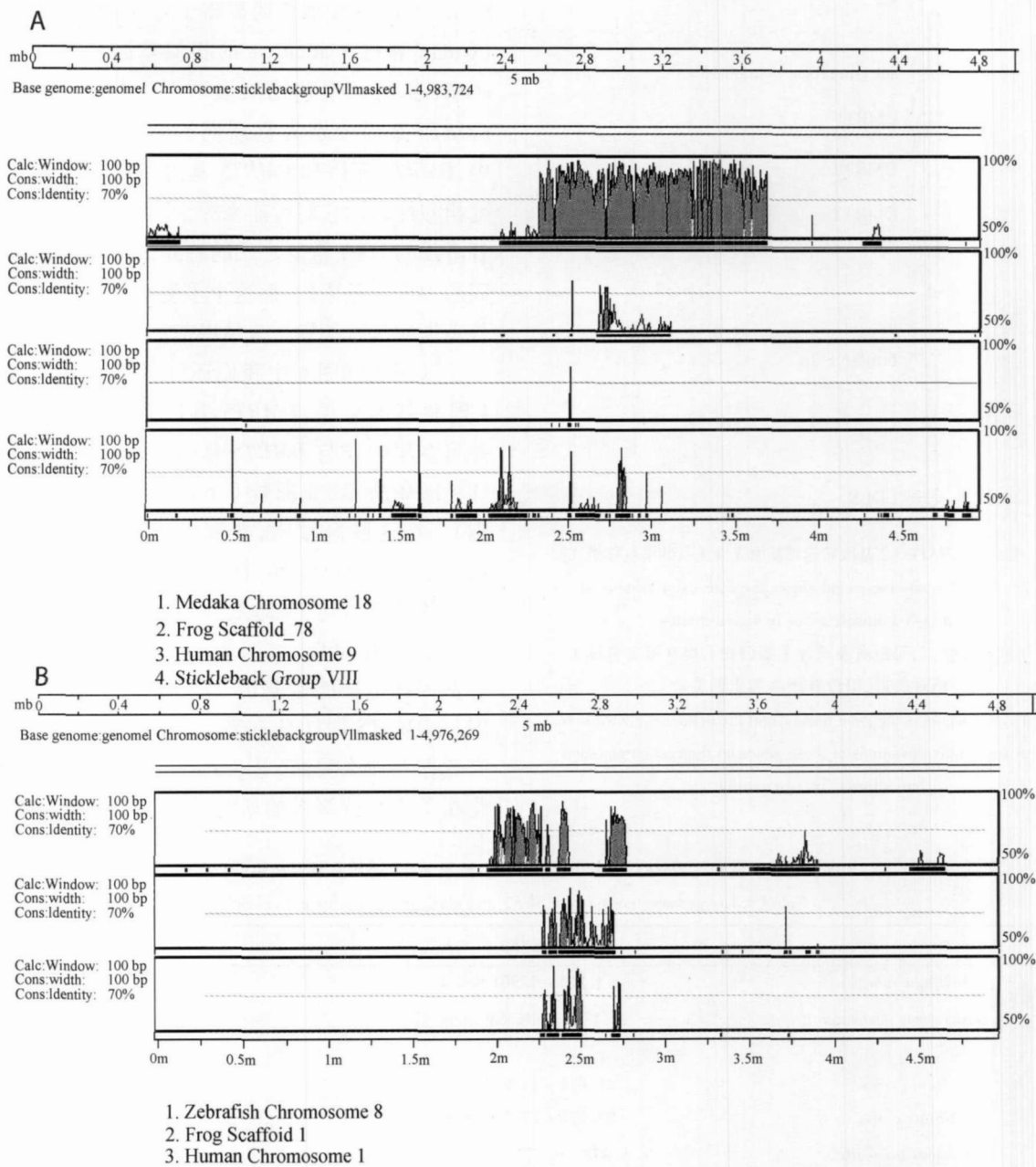


图 3 *DMRT4*基因 (A)和 *DMRT5* (B)基因所在染色体片段的序列相似性分析

Fig. 3 The sequence comparisons of *DMRT4* (A) and *DMRT5* (B) containing segments between different vertebrates' genomes 参与 *DMRT4*基因所在染色体片段序列分析的有: 三刺鱼 group VII染色体的 2.5—7.5M 区域、青鳉 18号染色体的 5—10M 区域、非洲蛙 scaffold_78 人 9号染色体的 20—25M 区域以及三刺鱼 group VIII染色体的 11.5—16.5M 区域; 参与 *DMRT5*基因所在染色体片段序列分析的有: 三刺鱼 group VIII染色体的 11.5—16.5M 区域、斑马鱼 8号染色体的 14.5—19.5M 区域、非洲蛙 scaffold_1 的 12.3—17.3M 区域以及人 1号染色体的 47.5—52.5M 区域。横坐标表示相应的基准染色体的位置, 纵坐标表示序列保守的水平

The segments compared here for *DMRT4* gene are follows the regions between 2.5M and 7.5M of stickle chromosome group VII, the regions between 5M and 10M of medaka chromosome 18, frog scaffold_78 the regions between 20M and 25M of human chromosome 9 and the regions between 11.5M and 16.5M of stickleback chromosome group VIII. The segments compared for *DMRT5* gene are follows the regions between 11.5M and 16.5M of stickle chromosome group VIII, the regions between 14.5M and 19.5M of zebrafish chromosome 8 the regions between 12.3M and 17.3M of frog scaffold_1 and the regions between 47.5M and 52.5M of human chromosome 1. The horizontal axes show the positions of base chromosome sequences and the level of conservation is displayed on the vertical axis

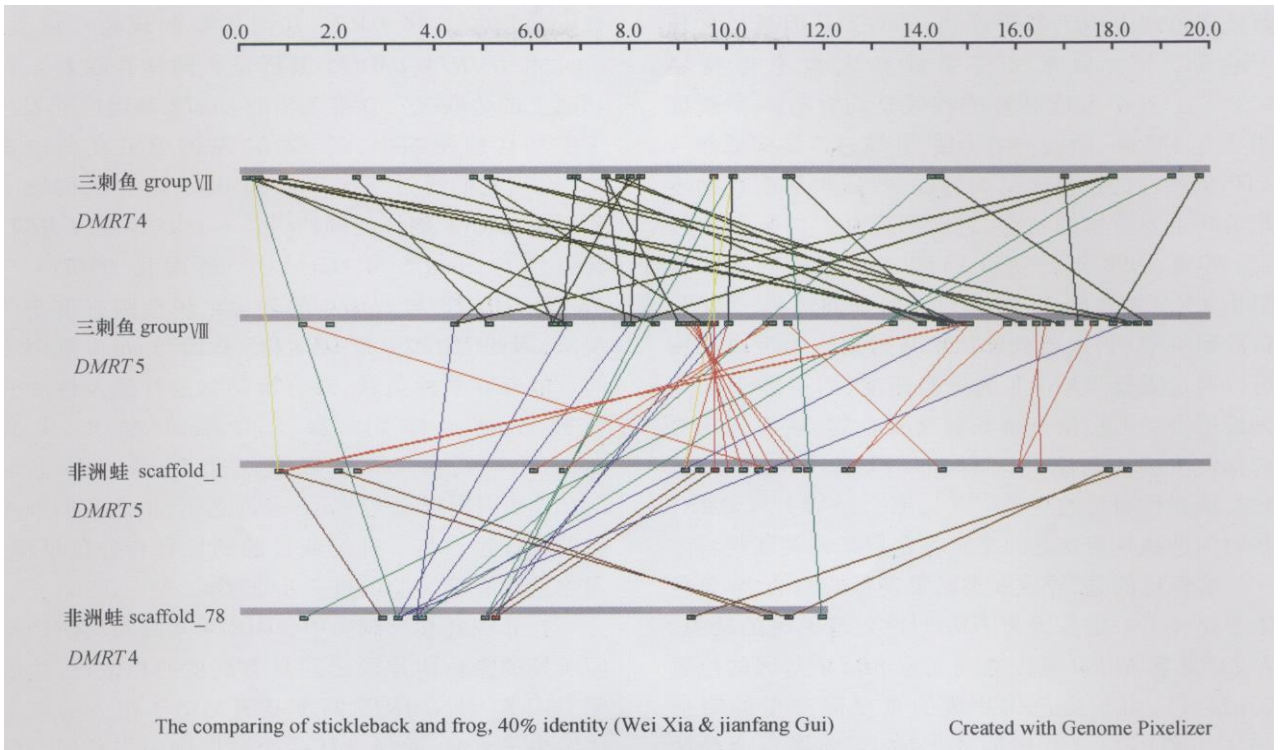


图 4 三刺鱼和非洲蛙的 *DMRT4*和 *DMRT5*所在染色体片段的同源基因比较

Fig 4 The duplication event of *DMRT4* and *DMRT5* containing chromosome segments in stickleback and frog

灰色粗线条表示染色体, 紧挨下面的小方框表示不同的重复基因 (红色小方框表示 *DMRT* 基因), 染色体左上角为所在染色体或 scaffold 的名称, 最上面为刻度。染色体之间编码蛋白序列相似程度超过 40% 的基因都被不同颜色的线条标出, 其中黑色、黄色和绿色线条分别连接三刺鱼 group VII 和 group VIII、非洲蛙 scaffold_1 以及 scaffold_78 的相似基因, 红色和蓝色线条分别连接三刺鱼 group VIII 和非洲蛙 scaffold_1 以及 scaffold_78 的相似基因, 棕色线条连接非洲蛙 scaffold_1 和 scaffold_78 的相似基因

Thick gray lines represent the chromosomes and the little boxes below the chromosomes represent different duplicated genes. Left of these chromosomes are their names and the scale is shown at the top of this figure. The lines in different color show the duplications between different chromosomes. All the genes at similarities above 40% are shown here

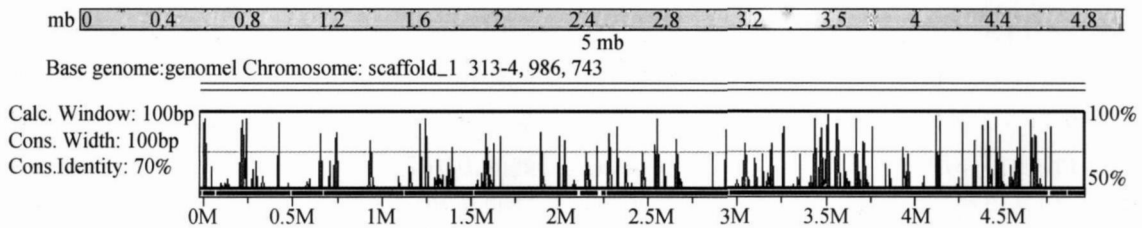


图 5 非洲蛙 *DMRT4*和 *DMRT5*基因所在染色体片段序列的比较 (纵横坐标同图 3)

Fig 5 The comparison of frog *DMRT4* and *DMRT5* containing chromosome segments (The axes are the same as Fig 3)

在过去不到十年的时间内, 各国研究人员分别对处于不同进化地位的多个物种的 *DMRT* 基因进行了大量的研究, 但绝大多数都是集中于对其表达模式的研究, 只有少量的工作是关于它们发挥作用的可能分子机制的研究^[20-26], 人们对在体内明确调控 *DMRT* 基因的表达和翻译, 以及受其调控的下游基因几乎还是一无所知, 同时对 *DMRT* 基因家族的演变历史也不是很清楚。因此本研究拟从基因组序列

比较入手, 分析该基因家族的演变过程, 并试图通过对基因连锁性的分析, 来鉴定潜在的与 *DMRT* 家族基因功能密切相关的基因。

研究表明, 鱼类和四足动物含有紧密连锁的 *DMRT1-DMRT3-DMRT2* 的基因组结构, 而且 *SNF2*, *LDLR* 及 *ANKRD15* 与该结构紧密连锁。对处于不同进化地位的多个物种的 *SNF2* 和 *DMRT* 的进一步连锁分析表明, *SNF2* 与几种存在或可能存在选择性

剪接现象的 *DMRT* 基因连锁。*SNF2* 基因最早发现于酵母^[27-29], 后来又在果蝇和人类中得到鉴定^[30-31]。这个家族共有的特征是均含有一个被证明具有 ATPase 活性的结构域, 并通过该结构域参与 ATP 依赖性的染色体结构的改变过程, 从而有助于相关转录复合物的结合, 进而调控基因的复制和转录。早在 1998 年就发现 *SNF2* 基因编码的蛋白会阻止 RNA 聚合酶 2 (RNAPII) 的延伸过程^[32]。最新的研究表明, 人类的 *hBRM* 基因编码的蛋白产物可以通过降低 RNAPII 的延伸速度, 而有助于在不是最适合的剪接位点聚集剪接复合物, 从而实现选择性剪接的调控^[33]。而 *DMRT1* 和 *DMRT2* 都已知有选择性剪接的现象^[34-36], 那么, *SNF2* 与 *DMRT* 基因的选择性剪接之间是否存在某种关联呢?

有意思的是, 导致人类精巢发育缺陷的 9p 染色体变化并不一定是因为 *DMRT1-3-2* 基因簇的缺失。在 *LDLR* 和 *hBRM* 基因之间靠近 *hBRM* 基因的位置 (9p24.1), 由于染色体片段的重复而产生的对应 9p21.1 和 9p23.3 片段的依次插入, 同样也会导致精巢发育的受阻^[37]。考虑到这种突变造成的表型与 *DMRT1-3-2* 基因簇缺失的表型非常相似, 我们猜测这种插入也可能导致了 *DMRT1-3-2* 基因簇的表达异常。由于插入的位置距离 *hBRM* 基因较近, 很可能是首先通过导致 *hBRM* 基因的表达异常, 从而进一步影响 *DMRT* 基因的表达。也就是说与 *DMRT1-3-2* 基因簇紧密连锁的 *hBRM* 基因, 极有可能在胎儿发育过程中起到调控 *DMRT1-3-2* 基因簇的表达, 从而促进精巢发育的作用。同时考虑到缺少选择性剪接的线虫 *mab3* 基因旁边没发现 *SNF2* 家族基因的存在, 我们推测 *SNF2* 家族基因可能是在调控 *DMRT1-3-2* 基因的转录乃至选择性的剪接方面发挥着重要的作用。但是由于缺少软骨鱼类以及头索动物的基因组数据, 目前尚不知道这些更低等物种的 *DMRT1-3* 基因在染色体上的分布状况, 因此我们暂时无法推测这种串联结构在鱼类和四足动物祖先中是否就已经形成了。

现有的研究表明, 脊椎动物的 *DMRT4* 基因除了在性腺表达之外, 还在胚胎发育时期嗅觉系统^[23-38-40] 的形成中发挥着重要的作用。关于 *DMRT5* 基因的研究则较少, 仅仅发现它的表达主要集中于脑部, 并在性腺中有微量表达^[40]。从鱼类 *DMRT4* 与 *DMRT5* 基因的比较结果中我们也同样发现了一个与上述类似的情况, 那就是与它们紧密连锁的 *elavL* 家族基因同样存在跟其他脊椎动物和无

脊椎动物的某些 *DMRT* 基因连锁的现象。那么, *elavL* 和 *DMRT4* *DMRT5* 基因是否同样存在着某种功能上的关联呢? 脊椎动物的 *elavL2* 基因广泛表达于脑和性腺组织中, 而 *elavL4* 基因则只在脑中表达^[41, 42], 它们通过三个串联的 RNA 结合结构域来实现对 mRNA 翻译的调控^[41, 43]。*DMRT4* 和 *DMRT5* 又分别与 *elavL2* 和 *elavL4* 连锁, 而且 *DMRT4* 与 *elavL2* *DMRT5* 与 *elavL4* 的表达组织范围有重叠的现象, 因此 *DMRT4* 和 *DMRT5* 基因与 *elavL* 基因之间可能存在某种关联, 对脊椎动物的神经及性腺发育和分化起作用。在无脊椎动物中, 果蝇 *DMRT11E*、线虫 *dmd-5* 以及海鞘 5q 染色体上的 *DMRT* 基因同样也与 *elavL* 基因连锁, 由此我们推测它们或许也参与了神经或性腺的发育与分化过程, 当然这还有待实验的进一步证实。

在非洲蛙和三刺鱼中, *DMRT4* 基因和 *DMRT5* 基因所处的染色体片段之间具有高度的相似性, 这表明 *DMRT4* 和 *DMRT5* 两个基因是由于祖先的染色体发生了重复, 而得以从一个共同的祖先基因分化而来。对其所含基因的线性分析可以清晰地看到, 相比 *DMRT4* 而言, 四足动物的 *DMRT5* 所在的染色体片段是更多地保留了祖先的基因的 (图 3B), 而 *DMRT4* 所在的染色体片段则发生了很多的改变, 乃至在鱼类和人类之间几乎找不到除 *DMRT* 和 *elav* 之外的任何同源基因, 这表明 *DMRT4* 和 *DMRT5* 所在的染色体片段经历了差异显著的进化历程。但是三刺鱼 *DMRT4* 和 *DMRT5* 分别所在的片段之间又有着多个同源基因, 只是我们的比对结果显示它们之间同源基因的相似性要显著低于三刺鱼和非洲蛙的 *DMRT5* 所在的片段, 这表明染色体片段重复事件发生的时间比较古老。同时我们发现在三刺鱼和非洲蛙的 *DMRT5* 所在染色体片段之间的同源基因, 与三刺鱼的 *DMRT4* 和 *DMRT5* 所在片段的同源基因不同 (反映在图 4 group VIII 上重叠的红线和黑线), 这表明鱼类和四足动物的相应染色体片段经历了有差异的选择。与四足动物不同, 鱼类 *DMRT4* 所在的染色体片段更加保守, 这样通过对比该区域附近基因的变化, 很有希望能够鉴定出一些鱼类特异、或四足动物特异的基因来。

现在已经完成的、正在进行的以及计划实施的基因组测序项目在不停地增长之中, 每天都有新的数据在互联网上发布出来, 充分利用这些宝贵的基因资源, 揭示基因进化的奥秘, 已成为一个新的学科生长点^[44]。因此, 本研究利用互联网上共享的基

基因组数据, 试图通过分析 *DMRT* 的基因组结构特征, 揭示 *DMRT* 基因家族的进化和起源, 为进一步研究 *DMRT* 基因的分子机制提供相关候选调控基因。通过本研究, 我们从基因组水平证明了真骨鱼类和四足动物相应 *DMRT* 基因的直系同源关系, 揭示了脊椎动物 *DMRT4* 和 *DMRT5* 基因的产生是由于它们的祖先经历了一次染色体重复事件而由同一基因演化而来的, 并且预测到了包括 *SNF2* 和 *elanL* 家族基因在内的多个可能与 *DMRT* 基因功能密切相关的基因, 为进一步研究 *DMRT* 基因和它们之间可能的调控机制提供了新的线索。

参考文献:

- [1] Raes J, Vandepoele K, Simillion C, *et al*. Investigating ancient duplication events in the Arabidopsis genome [J]. *J Struct Funct Genomics*, 2003, **3**: 117—129
- [2] Simillion C, Vandepoele K, Van Montagu M C, *et al*. The hidden duplication past of Arabidopsis thaliana [J]. *Proc Natl Acad Sci U S A*, 2002, **99**: 13627—13632
- [3] Taylor J S, Braasch J F, Rieckey T, *et al*. Genome duplication a trait shared by 22000 species of ray-finned fish [J]. *Genome Res*, 2003, **13**: 382—390
- [4] Christoffels A, Koh E G, Chia J M, *et al*. Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes [J]. *Mol Biol Evol*, 2004, **21**: 1146—1151
- [5] Jaillon O, Aury J M, Brunet F, *et al*. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype [J]. *Nature*, 2004, **431**: 946—957
- [6] Vandepoele K, De Vos W, Taylor J S, *et al*. Major events in the genome evolution of vertebrates: paraneome age and size differ considerably between ray-finned fishes and land vertebrates [J]. *Proc Natl Acad Sci U S A*, 2004, **101**: 1638—1643
- [7] Guo Y Q, Cheng H H, Gao S, *et al*. Phylogenetic tree and synteny of *DMRT* genes family of vertebrates [J]. *Acta Genetica Sinica*, 2004, **31**: 1103—1108 [郭一清, 程汉华, 高尚, 等. 脊椎动物 *DMRT* 基因家族的系统发生及同线性分析. 遗传学报, 2004, **31**: 1103—1108]
- [8] Volf J N, Zarkover D, Bardwell V J, *et al*. Evolutionary dynamics of the DM domain gene family in metazoans [J]. *J Mol Evol*, 2003, **57** Suppl 1: S241—249
- [9] Thomson J W, and DeSalle R. Gene family evolution and homology: genomics meets phylogenetics [J]. *Annu Rev Genomics Hum Genet*, 2000, **1**: 41—73
- [10] Xia W, Zhou L, Yao B, *et al*. Differential and spermatogenic cell-specific expression of *DMRT1* during sex reversal in protogynous hermaphroditic groupers [J]. *Mol Cell Endocrinol*, 2007, **263** (1—2): 156—172
- [11] Charles D. The origin of species [M]. University of Pennsylvania Press, 2006
- [12] Brunner B, Homung U, Shan Z, *et al*. Genomic organization and expression of the doublesex-related gene cluster in vertebrates and detection of putative regulatory regions for *DMRT1* [J]. *Genomics*, 2001, **77**: 8—17
- [13] Rice P, Longden I and Bleasby A. EMBOSS—the European Molecular Biology Open Software Suite [J]. *Trends Genet*, 2000, **16**: 276—277
- [14] Frazer K A, Pachter L, Poliakov A, *et al*. VISTA: computational tools for comparative genomics [J]. *Nucleic Acids Res*, 2004, **32**: W273—279
- [15] Sodergren E, Weinstock G M, Davidson E H, *et al*. The genome of the sea urchin *Strongylocentrotus purpuratus* [J]. *Science*, 2006, **314**: 941—952
- [16] Zhou L, Wang Y, Gui J F. Fish-specific genome duplication [J]. *Zoological Research*, 2006, **27**: 525—532 [周莉, 汪洋, 桂建芳. 鱼类特异的基因组复制. 动物学研究, 2006, **27**: 525—532]
- [17] Pevzner P, and Tesler G. Genome rearrangements in mammalian evolution: lessons from human and mouse genomes [J]. *Genome Res*, 2003, **13**: 37—45
- [18] Eichler E E, and Sankoff D. Structural dynamics of eukaryotic chromosome evolution [J]. *Science*, 2003, **301**: 793—797
- [19] Postlethwait J H, Woods I G, Ngo-Hazlett P, *et al*. Zebrafish comparative genomics and the origins of vertebrate chromosomes [J]. *Genome Res*, 2000, **10**: 1890—1902
- [20] Raymond C S, Murphy M W, O'Sullivan M G, *et al*. *Dmrt1* a gene related to worm and fly sexual regulators is required for mammalian testis differentiation [J]. *Genes Dev*, 2000, **14**: 2587—2595
- [21] Lei N, Heckert L L. Sp1 and Egr1 regulate transcription of the *Dmrt1* gene in Sertoli cells [J]. *Biol Reprod*, 2002, **66**: 675—684
- [22] Lei N, Heckert L L. Gata4 regulates testis expression of *Dmrt1* [J]. *Mol Cell Biol*, 2004, **24**: 377—388
- [23] Huang X, Hong C S, O'Donnell M, *et al*. The doublesex-related gene *XDmrt4* is required for neurogenesis in the olfactory system [J]. *Proc Natl Acad Sci U S A*, 2005, **102**: 11349—11354
- [24] Seo K W, Wang Y, Kokubo H, *et al*. Targeted disruption of the DM domain containing transcription factor *Dmrt2* reveals an essential role in somite patterning [J]. *Dev Biol*, 2006, **290**: 200—210
- [25] Bakuniene J, Bardwell V J, Zarkover D. Mice mutant in the DM domain gene *Dmrt4* are viable and fertile but have polyovular follicles [J]. *Mol Cell Biol*, 2006, **26** (23): 8984—8991
- [26] Meng A, Moore B, Tang H, *et al*. A Drosophila doublesex-related gene *terna* is involved in somitogenesis in vertebrates [J]. *Development*, 1999, **126**: 1259—1268
- [27] Surabhi R M, Daly L D, and Cattini P A. Evidence for evolutionary conservation of a physical linkage between the human BAF60b a subunit of SWI/SNF complex and thyroid hormone receptor interacting protein-1 genes on chromosome 17 [J]. *Genome*, 1999, **42**: 545—549
- [28] Hirschhorn J N, Brown S A, Clark C D, *et al*. Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure [J]. *Genes Dev*, 1992, **6**: 2288—2298
- [29] Tan K W, Deuring R, Scott M P, *et al*. Brahma a regulator of

- Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SW2 [J]. *Cell* 1992, **68**: 561—572
- [30] Papoulas O, Beek S J, Moseley S L, *et al*. The Drosophila trithorax group proteins BRM, ASH1 and ASH2 are subunits of distinct protein complexes [J]. *Development* 1998, **125**: 3955—3966
- [31] Okabe I, Bailey L C, Attree O, *et al*. Cloning of human and bovine homologs of SNF2/SW2: a global activator of transcription in yeast *S. cerevisiae* [J]. *Nucleic Acids Res*, 1992, **20**: 4649—4655
- [32] Liu M, Xie Z, and Price D H. A human RNA polymerase II transcription termination factor is a SW2/SNF2 family member [J]. *J Biol Chem*, 1998, **273**: 25541—25544
- [33] Batsche E, Yaniv M, and Muehhardt C. The human SW1/SNF subunit Brm is a regulator of alternative splicing [J]. *Nat Struct Mol Biol*, 2006, **13**: 22—29
- [34] Guo Y, Cheng H, Huang X, *et al*. Gene structure, multiple alternative splicing and expression in gonads of zebrafish *Dmrt1* [J]. *Biochem Biophys Res Commun*, 2005, **330**: 950—957
- [35] Huang X, Guo Y, Shui Y, *et al*. Multiple alternative splicing and differential expression of *dmrt1* during gonad transformation of the rice field eel [J]. *Biol Reprod*, 2005, **73**: 1017—1024
- [36] Ottolenghi C, Veitia R, Barbieri M, *et al*. The human doublesex-related gene *DMRT2* is homologous to a gene involved in somitogenesis and encodes a potential bicistronic transcript [J]. *Genomics* 2000, **64**: 179—186
- [37] Ion R, Tekli L, Chaussain J L, *et al*. Failure of testicular development associated with a rearrangement of 9p24, 1 proximal to the SNF2 gene [J]. *Hum Genet* 1998, **102**: 151—156
- [38] Kondo M, Froschauer A, Kitano A, *et al*. Molecular cloning and characterization of *DMRT* genes from the medaka *Oryzias latipes* and the platyfish *Xiphophorus maculatus* [J]. *Gene*, 2002, **295**: 213—222
- [39] Winkler C, Homung U, Kondo M, *et al*. Developmentally regulated and non-sex-specific expression of autosomal *dmrt* genes in embryos of the Medaka fish (*Oryzias latipes*) [J]. *Mech Dev*, 2004, **121**: 997—1005
- [40] Kin S, Kettlewell J R, Anderson R C, *et al*. Sexually dimorphic expression of multiple doublesex-related genes in the embryonic mouse gonad [J]. *Gene Expr Patterns* 2003, **3**: 77—82
- [41] Colegrove-Otero L J, Devaux A, and Standart N. The Xenopus ELAV protein ELB represses Vg1 mRNA translation during oogenesis [J]. *Mol Cell Biol*, 2005, **25**: 9028—9039
- [42] Good P J. A conserved family of elav-like genes in vertebrates [J]. *Proc Natl Acad Sci U S A*, 1995, **92**: 4557—4561
- [43] Devaux A, Colegrove-Otero L J, Standart N. Xenopus ELB but not ELA binds RNA as an oligomer: possible role of the linker [J]. *FEBS Lett*, 2006, **580**: 4947—4952
- [44] Bei J X, Zhang Y, Li W S, *et al*. Application of genome databases and bioinformatics in exploration for fish genes [J]. *Acta Hydrobiologica Sinica*, 2008, **32** (3): 387—392 [贝锦新, 张勇, 李文笙, 等. 鱼类基因数据库与生物信息学在鱼类基因开发上的应用. 水生生物学报, 2008, **32** (3): 387—392]

THE LINKAGE STRUCTURE AND PHYLOGENY OF *DMRT* GENE FAMILY IN THE TELEOST FISH

XIA Wei^{1,2}, ZHOU Li¹ and GU I Jian-Fang¹

(1. State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology,

Chinese Academy of Sciences, Wuhan 430072; 2. Graduate School of the Chinese Academy of Sciences, Beijing 100039)

Abstract The DM domain gene families encode putative transcription factors related to the sexual regulators *dsx* from *Drosophila melanogaster* and *mab-3* from *Caenorhabditis elegans*. *DMRT* genes were first identified as genes involved in sexual development in very distant metazoan phyla but some of them were later shown to be involved in the somite or neuro system development in the embryo development. So different *DMRT* genes must play distinct roles, just the exact function mechanisms of the great majority of them, especially those of vertebrates *DMRT* genes remain unclear. For example we know really little about their upstream regulators and downstream targets.

DM domain genes underwent frequent independent events of gene duplication during the course of evolution and the number of *DMRT* genes differs between phyla even within vertebrate lineage. Thanks to the existing genomic data of different species we can compare the adjacent regions of *DMRT* genes (from *DMRT1* to *DMRT5*) from different teleost fishes. In this study, we fully utilize the genomic data that we can get free from the website and use the well established method in comparing large genome sequences to analyze the relationship between different *DMRT* containing segments from different species genomes.

Firstly we compared the *DMRT1-3-2* cluster containing chromosome segments from different fish species with those from tetrapods and proved the fact that descended from one common chromosome segment before the divergence of ray-

finned fish and tetrapods. When we further performed the linkage analysis of this region, we also identified three genes of *SNF2*, *LDLR* and *ANKRD15*, which were conserved in all the vertebrates we analyzed here. And the most interesting thing was that the *SNF2* gene was closely linked to some *DMRT* genes across very distant metazoan phyla. The comparative sequence analysis also proved the orthologous relationships between *DMRT4* or *DMRT5* genes of different vertebrates, and the results uncovered an ancient chromosome duplication (or segmental duplication) event happened to the common ancestor of vertebrates *DMRT4* and *DMRT5* genes. Furthermore, we identified lots of genes including *ekvL* genes that might have functions of close relationships with *DMRT4* and *DMRT5* by investigating their linked genes in different vertebrates, which further provided clues for in depth studies about the functions of *DMRT* family.

Key words Genome Link Evolution Gene duplication Sequence comparison