*Hans Journal of Biomedicine* 生物医学, 2011, 1, 6-16
doi:10.4236/hjbm.2011.11002 Published Online July 2011 (http://www.hanspub.org/journal/hjbm/)

# Genome-Wide Survey, Identification and Preliminary Analysis of Xenopus Laevis Bhlh Transcription Factors

**Wuyi Liu**[1,2]
[1] Department of Biology Science, Fuyang Normal College, Fuyang
[2]Department of Scientific Research, Fuyang Normal College, Fuyang
Email: lwycau@yahoo.com.cn
Received: May 15th, 2011; revised: Jun. 25th, 2011; accepted: Jun. 30th, 2011.

**Abstract:** The basic helix-loop-helix (bHLH) transcription factors play essential roles in the regulation of eukaryotic growth and development and gene transcription. In this study, we conducted a genome-wide survey using the *Xenopus Laevis* ongoing genome project databases, and identified 98 bHLH sequences in *Xenopus Laevis* genome. Phylogenetic analyses revealed those bHLH genes belong to 32 families in the super-groups (A-F) in this research. Gene Ontology (GO) enrichment statistics showed 42 significant GO annotations counted in frequency. Statistical analysis of the Gene Ontology annotations showed that these 98 bHLH proteins tend to be related to transcription regulator activity (GO: 0030528), regulation of transcription (GO: 0045449), DNA binding (GO: 0003677), transcription (GO: 0006350), DNA-dependent regulation of transcription (GO: 0006355), expected from the common GO categories of transcriptional factors. A number of bHLH genes play regulation significant role in special development or physiology processes, such as muscle organ development and eye development. This preliminary study provides useful information for further researches on *Xenopus Laevis*.

**Keywords:** Transcription Factor; Gene Ontology; Genome-Wide Analysis

# 非洲爪蟾的碱性螺旋－环－螺旋转录因子的鉴定与初步分析

**刘武艺** [1,2]
[1] 阜阳师范学院生物系，阜阳
[2] 阜阳师范学院科研处，阜阳
Email: lwycau@yahoo.com.cn
收稿日期：2011 年 5 月 15 日；修回日期：2011 年 6 月 25 日；录用日期：2011 年 6 月 30 日

**摘　要：**碱性螺旋－环－螺旋(bHLH)转录因子在真核生物的生长发育相关的基因表达调控过程中发挥着重要的作用。本文根据现有非洲爪蟾基因组数据，利用生物信息学方法初步鉴定爪蟾的 bHLH 基因，收集了其结构、家族分类和 GO 功能富集等信息，进行了初步分析。结果，在非洲爪蟾基因组数据库中共发现 98 个 bHLH 转录因子，它们可以分别归到 6 大组(A-F)的 32 个亚家族中。通过基因本体论(GO)的富集分布统计发现有 42 个显著富集分布的 GO 注释语句，其中转录调控活性(GO: 0030528)、转录调控(GO: 0045449)、DNA 结合(GO: 0003677)、转录(GO: 0006350)和 DNA 依赖的转录调控(GO: 0006355)等出现的频率很高，表明这些 GO 术语是爪蟾 bHLH 基因的共同功能；此外，许多爪蟾 bHLH 基因在一些重要的发育或生理过程(如肌肉器官和眼的发育)中发挥着重要的调控作用。这些研究结果将有助于进一步的研究。

**关键词：**转录因子；基因本体论；基因组分析

## 1. 引言

转录因子(transcription factor)，又称反式作用因子，指能够与真核基因顺式作用元件(*cis* acting element)发生特异的相互作用并对转录有激活或抑制作用的 DNA 结合蛋白，转录因子调控复杂的蛋白间互作网络[1]。典型的转录因子含有 DNA 结合区、转

录调控区、寡聚化位点及核定位信号区等[1-5]。有关转录因子结构和功能的研究是动植物分子生物学研究的前沿领域，转录因子因其含有 DNA 结合蛋白的不同可以划分为不同的基因家族[1-3]。碱性螺旋－环－螺旋(basic helix-loop-helix，bHLH)转录因子是目前最大的转录因子家族之一，并被公认为在细胞增殖与分化、肌肉形成、神经元、肠和血、性别决定等遗传发育过程中具有重要作用[4-8]，许多课题组对 bHLH 转录因子展开了研究。最早报道的是鼠转录因子 E12 和 E47[9]，后来的研究将动物 bHLH 转录因子划分为 6 大类，这些大类又细分为 45 个亚类或蛋白家族[4,6,10,11]。现在，动物的 bHLH 家族成员的分类、进化及功能分析现已积累大量资料，且基本清楚各家族组成及成员的功能。由于生物物种基因组测序和全基因组草图绘制工作的陆续完成，越来越多的转录因子被分析和鉴定出来，这就为从整体上研究某物种转录因子的功能和进化等重要问题提供了可能。因此，从全基因组角度研究某一类型的转录因子或调控因子具有重要的意义。现今，动物 bHLH 转录因子家族已经在人、小鼠、大鼠、鸡、蚕、蜜蜂等许多物种的基因组中被鉴定和研究[4-17]。但在非洲爪蟾(*Xenopus laevis*)基因组中尚未进行有关的研究。

非洲爪蟾是较早用于生物医学的重要模式动物之一[19]。本研究以 Atchley et al. (1999)的 bHLH 转录因子分类原则[10]，Ledent et al. (2001, 2002)定义的 45 个 bHLH 代表性基序(domains)和 118 个人 bHLH 基因序列为鉴定标准[6,11]，从非洲爪蟾的基因组数据库鉴定出 98 个 bHLH 转录因子，进行基因本体论(Gene Ontology，GO)富集分析，以期分析和了解这些 bHLH 基因的功能信息。

## 2. 材料与方法

### 2.1. BLAST 搜索和 bHLH 转录因子的鉴定

根据 Atchley et al. (1999)的 bHLH 转录因子分类原则和 Ledent et al. (2001, 2002)定义的 45 个代表性 bHLH 基序[6,11]进行 TBLASTN 和 BLASTP，搜索候选 bHLH 基序。其中，每条序列都被用于对 NCBI 的非洲爪蟾的基因组数据库[19]进行反复搜索(http://www.ncbi.nlm.nih.gov/genome/guide/frog/)，搜索的严谨值设为 $E < 10$，以获取所有可能的 bHLH 序列。同时，我们也检索蛙蟾数据库 Xenbase[20]，最后根据 scaffold 或基因克隆(genomic clone)的编号、编码区、基因和蛋白获取号、序列比对的结果等信息，去除冗余序列，得到最终采用的 bHLH 因子序列。

### 2.2. 序列比对和基序比较

通过 BLAST 搜索得到的爪蟾蛋白质序列，用 ClustalX 2.0[21]进行比对，接着用 GeneDoc 2.6[12]做保守序列的分析和比较。

### 2.3. 基因本体(Gene ontology，GO)注释的富集分布分析

利用 DAVID 生物信息工具[22,23]做 GO 注释的富集分析，富集分布的显著性 *P* 值和假阳性率(False positive rate，FDR)均被控制在 0.05 以下。

### 2.4. BHLH 同源基因的系统发育分析

系统发育分析(Phylogenetic analyses)用最大似然估计软件 PHYML 2.4.4[24]和贝叶斯推断软件 Mrbayes 3.12[25]进行系统发育树推断。其中，贝叶斯推断采用两个独立的马科夫链进行推断(进行 14,000,000 步马科夫链抽样，抽样频率为每代抽样 100 个)，取 50%一致树作为最终的系统发育树。

## 3. 结果与讨论

### 3.1. 爪蟾 bHLH 转录因子的鉴定

利用上述 bHLH 代表性基序、TBLASTN 和 BLASTP 算法及系统发育分析，搜索鉴定得到了 98 条爪蟾的 bHLH 序列(图 1、表 1)。图 1 所示的 bHLH 转录因子名称通过与人的同源序列(homolog)的系统发育树分析得到。若一条人的 bHLH 因子拥有两个以上爪蟾同源序列，我们将之分别标注为 a、b、c，或 1、2、3 等名称。例如，人的 Hath5 和 Hath4a 在爪蟾基因组中发现有两个同源序列，那么爪蟾的相应同源序列就会被命名为 Xath5a、Xath5b，和 Xath4a1、Xath4a2。本研究发现，爪蟾的 98 个 bHLH 转录因子分别有 39、22、13、5、16 和 3 个因子可以被归类到 6 个高阶组(high-order group)A-F 的 32 个小家族；而某些小家族，如 Mist、Beta3、Oligo、Net、Delilah、MyoRb、PTFa、PTFb、AP4、MLX 和 TF4 等家族的成员没有

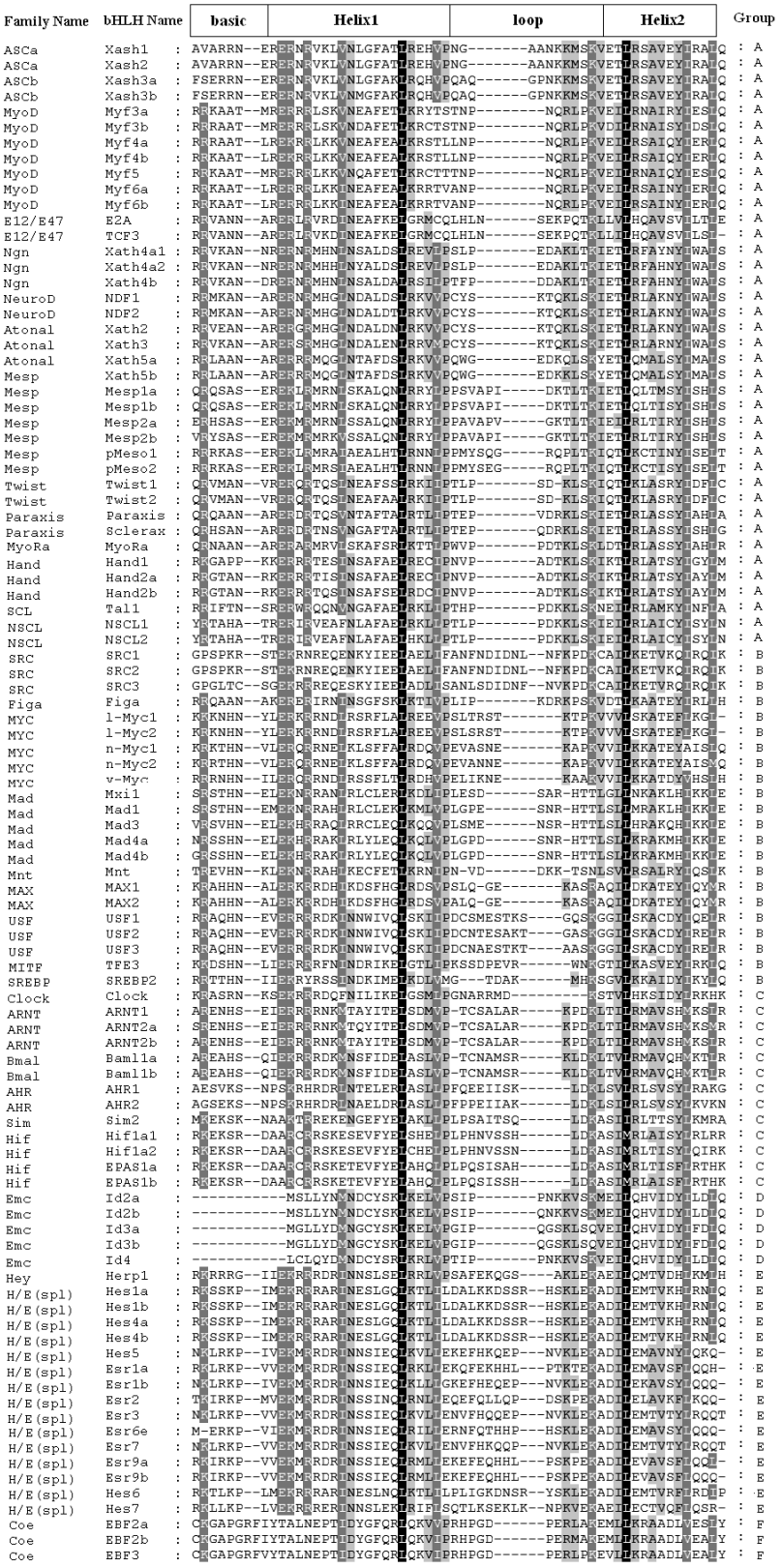



**Figure 1. Alignment of 98 Xenopus laevis bHLH domains (conserved sites are shaded and highly conserved sites are shaded in black)**

**图 1. 98 个非洲爪蟾 bHLH 基因基序的蛋白质序列信息(保守位点被涂上了阴影，其中高度保守的位点涂布了黑色的阴影)**

**Table 1. 98 bHLH genes in phylogenetic analysis and protein identification information**
**表 1. 98 个 bHLH 基因的系统发育分析和蛋白质鉴定信息**

| bHLH 基因家族 | 基因名称 | 各因子与人同源序列的系统发育分析信息 | | | 蛋白质获取号 |
|---|---|---|---|---|---|
| | | 同源基因 | ML自展值(%) | BI 后验概率(%) | |
| ASCa | *Xash*1 | *Hash*1 | *n/m** | 80 | NP_001079247.1 |
| ASCa | *Xash*2 | *Hash*2 | 96 | 67 | NP_001085994.1 |
| ASCb | *Xash*3*a* | *Hash*3*a*<br>*Hash*3*b*<br>*Hash*3*c* | *n/m* | *n/m* | NP_001079106.1 |
| ASCb | *Xash*3*b* | *Hash*3*a*<br>*Hash*3*b*<br>*Hash*3*c* | *n/m* | *n/m* | NP_001079125.1 |
| MyoD | *Myf*3*a* | *Myf*3 | 93 | 83 | NP_001079366.1 |
| MyoD | *Myf*3*b* | *Myf*3 | 93 | 83 | NP_001081292.1 |
| MyoD | *Myf*4*a* | *Myf*4 | 82 | 99 | NP_001079326.1 |
| MyoD | *Myf*4*b* | *Myf*4 | 88 | 99 | NP_001079199.1 |
| MyoD | *Myf*5 | *Myf*5 | 51 | 59 | NP_001095249.1 |
| MyoD | *Myf*6*a* | *Myf*6 | *n/m** | 94 | NP_001081477.1 |
| MyoD | *Myf*6*b* | *Myf*6 | *n/m** | 94 | NP_001088572.1 |
| E12/E47 | *E*2*A* | *E*2*A* | 82 | *n/m* | NP_001080409.1 |
| E12/E47 | *TCF*3 | *TCF*3 | *n/m** | 88 | NP_001079668.1 |
| Ngn | *Xath*4*a*1 | *Hath*4*a* | 97 | 100 | NP_001081802.1 |
| Ngn | *Xath*4*a*2 | *Hath*4*a* | 97 | 100 | NP_001081804.1 |
| Ngn | *Xath*4*b* | *Hath*4*b* | 83 | 91 | NP_001128257.1 |
| NeuroD | *NDF*1 | *NDF*1 | 82 | 97 | NP_001079263.1 |
| NeuroD | *NDF*2 | *NDF*2 | 82 | 97 | NP_001085596.1 |
| Atonal | *Xath*2 | *Hath*2 | *n/m** | 79 | NP_001079218.1 |
| Atonal | *Xath*3 | *Hath*3 | 97 | 99 | NP_001081213.1 |
| Atonal | *Xath*5*a* | *Hath*5 | 95 | 100 | NP_001079289.1 |
| Atonal | *Xath*5*b* | *Hath*5 | 95 | 100 | NP_001079290.1 |
| Mesp | *Mesp*1*a* | *Mesp*1<br>*Mesp*2<br>*pMesp*1 | *n/m* | *n/m* | NP_001128698.1 |
| Mesp | *Mesp*1*b* | *Mesp*1<br>*Mesp*2<br>*pMesp*1 | *n/m* | *n/m* | NP_001091431.1 |
| Mesp | *Mesp*2*a* | *Mesp*1<br>*Mesp*2<br>*pMesp*1 | *n/m* | *n/m* | NP_001079050.1 |
| Mesp | *Mesp*2*b* | *Mesp*1<br>*Mesp*2<br>*pMesp*1 | *n/m* | *n/m* | NP_001081641.1 |
| Mesp | *pMeso*1 | *pMesp*1 | 99 | 100 | NP_001083813.1 |
| Mesp | *pMeso*2 | *pMesp*1 | 99 | 100 | NP_001136111.1 |
| Twist | *Twist*1 | *Twist*1<br>*Twist*2 | 98 | *n/m* | NP_001079352.1 |
| Twist | *Twist*2 | *Twist*1<br>*Twist*2 | 98 | *n/m* | NP_001091211.1 |
| Paraxis | *Paraxis* | *Paraxis* | 62 | 77 | NP_001087941.1 |
| Paraxis | *Sclerax* | *Sclerax* | 80 | 100 | NP_001092152.1 |

| | | | | | |
|---|---|---|---|---|---|
| MyoRa | *MyoRa* | *MyoRa*1<br>*MyoRa*2 | 82 | 100 | NP_001085957.1 |
| Hand | *Hand*1 | *Hand*1 | 92 | 100 | NP_001079128.1 |
| Hand | *Hand*2*a* | *Hand*2 | 98 | 100 | NP_001079108.1 |
| Hand | *Hand*2*b* | *Hand*2 | 98 | 100 | NP_001107665.1 |
| SCL | *Tal*1 | *Tal*1 | *n/m** | 56 | NP_001081746.1 |
| NSCL | *NSCL*1 | *NSCL*1 | *n/m** | 100 | NP_001081852.1 |
| NSCL | *NSCL*2 | *NSCL*2 | 89 | 100 | NP_001088421.1 |
| SRC | *SRC*1 | *SRC*1 | 98 | 100 | NP_001154867.1 |
| SRC | *SRC*2 | *SRC*2 | 92 | 100 | NP_001081139.1 |
| SRC | *SRC*3 | *SRC*3 | 78 | 99 | NP_001081732.1 |
| Figα | *Figα* | *Figα* | 87 | 100 | NP_001088667.1 |
| MYC | *l-Myc*1 | *L-Myc*1 | 99 | 100 | NP_001081340.1 |
| MYC | *l-Myc*2 | *L-Myc*2 | 54 | 100 | NP_001079460.1 |
| MYC | *n-Myc*1 | *n-Myc* | 71 | 99 | NP_001079365.1 |
| MYC | *n-Myc*2 | *n-Myc* | 71 | 99 | NP_001084122.1 |
| MYC | *v-Myc* | *v-Myc* | 88 | 100 | NP_001080349.1 |
| Mad | *Mxi*1 | *Mxi*1 | *n/m* | 97 | NP_001089170.1 |
| Mad | *Mad*1 | *Mad1a* | *n/m** | 71 | NP_001090200.1 |
| Mad | *Mad*3 | *Mad*3 | 99 | 100 | NP_001090188.1 |
| Mad | *Mad4a* | *Mad*4 | 84 | 97 | NP_001079167.1 |
| Mad | *Mad4b* | *Mad*4 | 84 | 97 | NP_001084456.1 |
| Mnt | *Mnt* | *Mnt* | 72 | 99 | NP_001089310.1 |
| MAX | *MAX*1 | *MAX* | 86 | 100 | NP_001079118.1 |
| MAX | *MAX*2 | *MAX* | 86 | 100 | NP_001089042.1 |
| USF | *USF*1 | *USF*1 | 98 | 100 | NP_001089471.1 |
| USF | *USF*2 | *USF*2 | 99 | 100 | NP_001088134.1 |
| USF | *USF*3 | *USF*3 | 99 | 100 | NP_001088700.1 |
| MITF | *TFE*3 | *TFE*3 | 91 | 88 | NP_001088215.1 |
| SREBP | *SREBP*2 | *SREBP*2 | 82 | 97 | NP_001085554.1 |
| Clock | *Clock* | *Clock* | 100 | 100 | NP_001083854.1 |
| ARNT | *ARNT*1 | *ARNT*1 | *n/m** | 100 | NP_001082130.1 |
| ARNT | *ARNT*2*a* | *ARNT*2 | 99 | 100 | NP_001080540.1 |
| ARNT | *ARNT*2*b* | *ARNT*2 | 99 | 100 | NP_001083622.1 |
| Bmal | *Baml*1*a* | *Bmal*1 | *n/m** | 59 | NP_001089024.1 |
| Bmal | *Baml*1*b* | *Bmal*1 | *n/m** | 59 | NP_001089031.1 |
| AHR | *AHR*1 | *AHR*1 | 91 | 100 | NP_001082693.1 |
| AHR | *AHR*2 | *AHR*2 | 94 | 100 | NP_001121349.1 |
| Sim | *Sim*2 | *Sim*2 | 82 | 97 | NP_001079101.1 |
| HIF | *Hif1α*1 | *Hif*1*α* | 99 | 54 | NP_001086426.1 |
| HIF | *Hif1α*2 | *Hif*1*α* | 99 | 54 | NP_001080449.1 |
| HIF | *EPAS*1*a* | *EPAS*1 | 87 | 92 | NP_001085564.1 |
| HIF | *EPAS*1*b* | *EPAS*1 | 87 | 92 | NP_001085718.1 |
| Emc | *Id*2*a* | *Id*2 | 75 | 69 | NP_001087639.1 |

| | | | | | |
|---|---|---|---|---|---|
| Emc | *Id2b* | *Id*2 | 75 | 69 | NP_001081902.1 |
| Emc | *Id3a* | *Id*3 | 96 | 100 | NP_001079535.1 |
| Emc | *Id3b* | *Id*3 | 96 | 100 | NP_001079757.1 |
| Emc | *Id*4 | *Id*4 | 89 | 74 | NP_001080704.1 |
| Hey | *Herp*1 | *Herp*1 | 87 | 97 | NP_001083926.1 |
| H/E (spl) | *Hes*1*a* | *Hes*1 | 62 | 52 | NP_001081396.1 |
| H/E (spl) | *Hes*1*b* | *Hes*1 | 62 | 52 | NP_001079386.1 |
| H/E (spl) | *Hes*4*a* | *Hes*4 | 85 | 98 | NP_001082574.1 |
| H/E (spl) | *Hes*4*b* | *Hes*4 | 85 | 98 | NP_001082161.1 |
| H/E (spl) | *Hes*5 | *Hes*5 | 92 | 100 | NP_001079464.1 |
| H/E (spl) | *Esr*1*a* | *Hes*5 | 92 | 100 | NP_001079236.1 |
| H/E (spl) | *Esr*1*b* | *Hes*5 | 92 | 68 | NP_001089096.1 |
| H/E (spl) | *Esr*2 | *Hes*5 | 92 | 100 | NP_001082163.1 |
| H/E (spl) | *Esr*3 | *Hes*5 | 92 | 95 | NP_001089095.1 |
| H/E (spl) | *Esr6e* | *Hes*5 | 92 | 100 | NP_001081972.1 |
| H/E (spl) | *Esr*7 | *Hes*5 | 92 | 95 | NP_001081974.1 |
| H/E (spl) | *Esr*9*a* | *Hes*5 | 92 | 100 | NP_001081706.1 |
| H/E (spl) | *Esr*9*b* | *Hes*5 | 92 | 100 | NP_001089097.1 |
| H/E (spl) | *Hes*6 | *Hes*6 | *n/m** | 100 | NP_001116354.1 |
| H/E (spl) | *Hes*7 | *Hes*7 | 88 | 91 | NP_001082175.1 |
| Coe | *EBF*2*a* | *EBF*2 | *n/m** | 84 | NP_001079146.1 |
| Coe | *EBF*2*b* | *EBF*2 | *n/m** | 84 | NP_001079147.1 |
| Coe | *EBF*3 | *EBF*3 | *n/m** | 100 | NP_001083801.1 |

注释：最大似然自展值指由最大似然方法(ML)构建进化树得到的各分支上的自展值(Bootstrap Value)；贝叶斯后验概率指根据贝叶斯推断方法(BI)构建进化树得到的各个分支上的后验概率；问号表示不匹配；*n/m*标记表示某个基因不能与单一的同源基因形成独立的分支，但是可以与一群同源基因形成一个分支；*n/m** 标记表示某个基因与同源基因形成独立的分支时自展值小于50%。

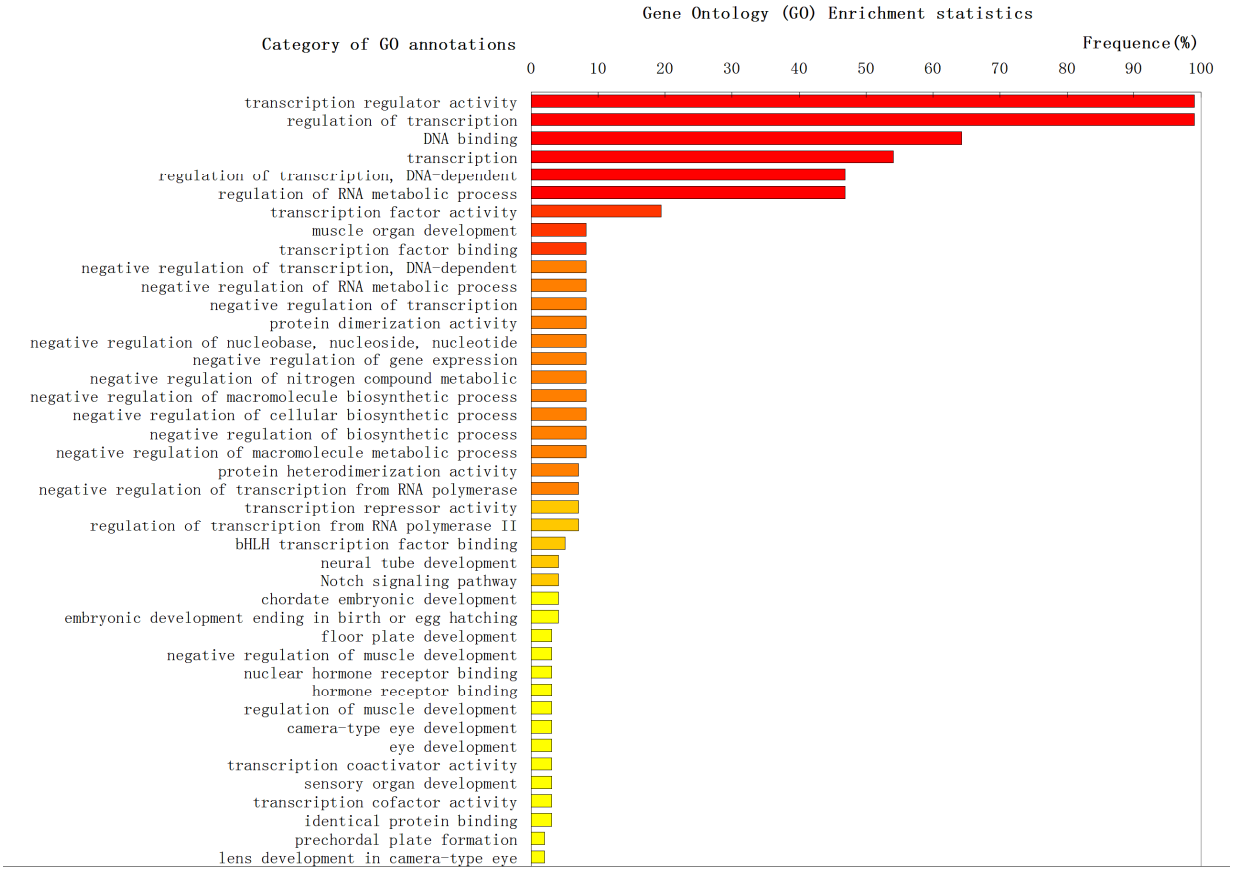
被发现(表 1)。本研究中，我们共发现了 16 个预测的 bHLH 基因，即 NP_001085994.1、NP_001088572.1、NP_001079668.1、NP_001085596.1、NP_001091211.1、NP_001088421.1、NP_001154867.1、NP_001088667.1、NP_001089471.1、NP_001088134.1、NP_001088700.1、NP_001089031.1、NP_001085564.1、NP_001085718.1、NP_001087639.1 和 NP_001079757.1，这些基因都是通过生物信息学推断或预测的新转录因子，为在蟾蜍中新的人类 bHLH 蛋白的同源蛋白，其在数据库中尚未有详细的注释。此外，本研究还验证了 3 个基因、纠正了 2 个可能被错误注释的基因，即 Paraxis (NP_001087941.1)、Id3b (NP_001079757.1)、Hes5 (NP_001079464.1)和 Mesp2a (NP_001079050.1，原名 Thylacine1)、Mesp2b (NP_001081641.1，原名 Thylacine2)。本研究得到的结果可以作为基因组数据库中注释信息的有益补充。

需要注意的是，本研究的数据可能没有获取所有的 bHLH 基因信息，或者得到的基因中存在假基因。因为，一方面非洲爪蟾(*Xenopus laevis*)的基因组测序工作尚未完成，另一方面非洲爪蟾在进化早期经历了“染色体多倍化”(Tetraploidization)，其基因组大致相当西方爪蟾(*Xenopus tropicalis*)的两倍[26]，例如 *Myf*3、*Myf*4 和 *Myf*6 等基因就分别发现了两个拷贝。但非洲爪蟾基因组的大部分基因都是很保守的，多倍化的染色体对其基因功能突变的影响相对比较小[26]。总之，获取的数据对研究结果没有太大的影响。

### 3.2. BHLH 转录因子基因本体论(GO)富集分布和功能注释信息分析

一般而言，DNA 结合活性和蛋白质聚合活性、转录共




注释：所有的GO语句均来自国际基因本体数据库(http://www.geneontology.org)。

**Figure 2. Forty-two significant GO annotations counts plotted by frequency**
**图 2. 42 个统计显著的 GO 注释(GO Aannotation)出现频次的柱状图统计**

录共激活是 bHLH 类因子的主要功能活动。但是，除了这些转录因子共有的功能之外，bHLH 转录因子还有其自身特殊的功能活性。为了进一步探讨爪蟾 bHLH 转录因子家族的整体功能特点，我们收集了这 98 个 bHLH 因子的基因本体论(GO)的功能注释信息。其中，42 个超几何分布统计检验显著($P < 0.05$)的 GO 注释语句显示在图 2 中，这些 GO 语句表示了一些重要的生物学过程、分子功能和信号通路(Pathway)信息，如转录调控活性(GO: 0030528)、转录调控(GO: 0045449)、DNA 结合(GO: 0003677)、转录(GO: 0006350)和 DNA 依赖的转录调控(GO: 0006355)等出现的频率很高，表明这些 GO 注释语句是爪蟾 bHLH 基因常见的功能。

同时，爪蟾 bHLH 转录因子家族的 GO 注释语句显示，除了共同拥有的功能注释信息之外，一些重要的发育过程或生理过程，如肌肉器官发育、神经管发育、胚胎脊索发育、血小板、照相机型眼和普通眼的发育、核内激素受体结合(nuclear hormone receptor binding)和激素受体结合、感官发育及 Notch 信号通路(notch signaling pathway)等出现的频率也较高。另外，6 大高阶组亦具有各个组内其自身的 GO 富集分布特点。图 2 中显示的均为 GO 超几何分布中统计上显著富集的功能注释($P < 0.05$，FDR < 0.05)。

### 3.3. 脊椎动物和无脊椎动物中 bHLH 基因数目和分布特点的比较及 Hes 基因家族的进化

为了解蟾蜍类与其他动物基因组 bHLH 转录因子的差异，我们比较了脊椎动物和无脊椎动物基因组中 bHLH 基因数目及其分布。脊椎动物的 bHLH 基因数明显地比无脊椎动物的要多(表 2)。有些基因家族，如 E12/E47、NeuroD、Atonal、Mesp、Twist、Paraxis、SCL、SRC、Myc、Mad、MITF、HIF、Emc、Hey 和 Coe 等在脊椎动物是多基因家族，而在无脊椎动物中

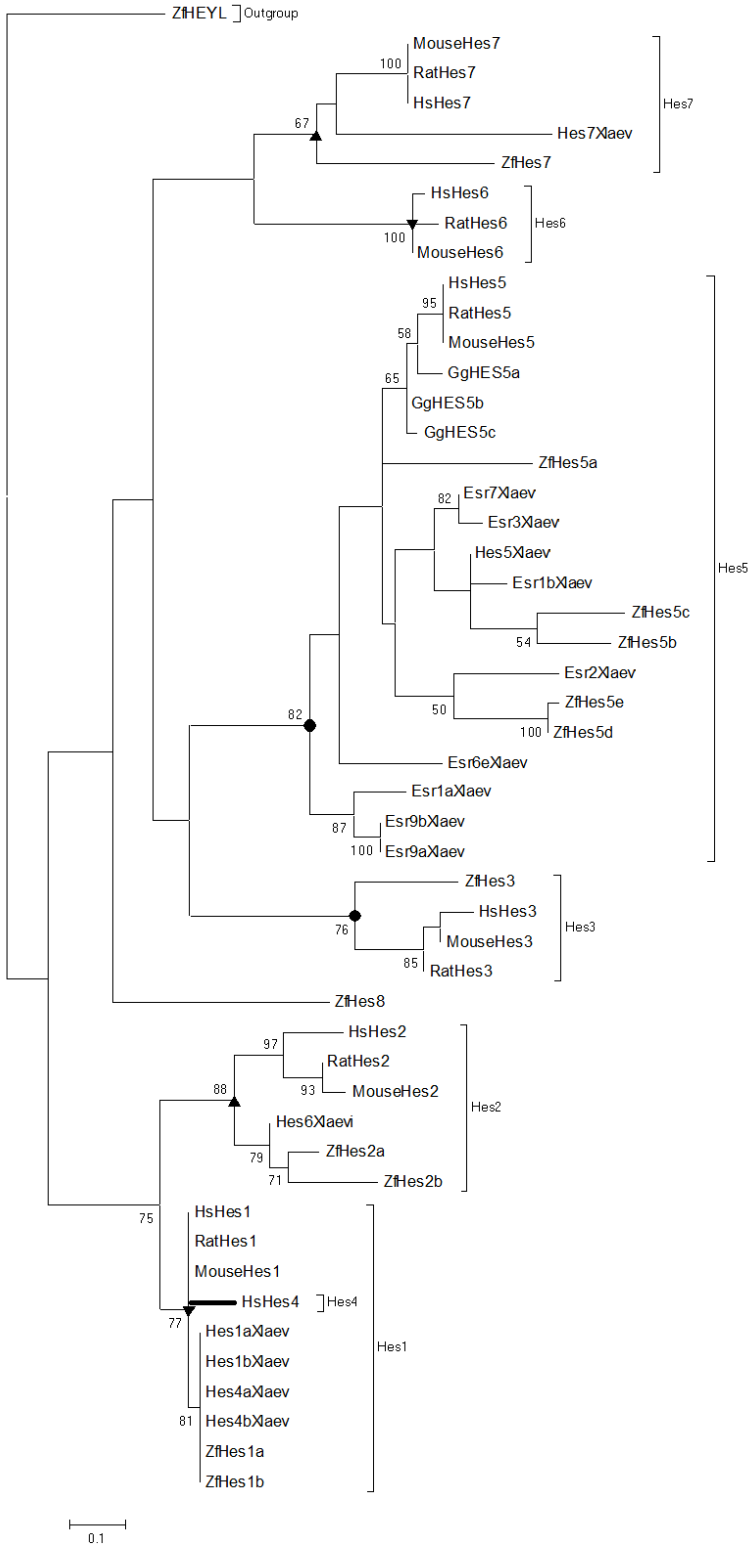
 

**Table 2. Comparing the number of bHLH transcription factors found among vertebrate and invertebrate species**
**表 2. 脊椎动物和无脊椎动物中 bHLH 基因数目和分布特点之比较**

| Family | Group | Drosophila | Lancelet | Giant owl limpet | *Xenopus Laevis* | Chicken | Zebrafish | Rat | Mouse |
|---|---|---|---|---|---|---|---|---|---|
| ASCa | A | 4 | 3 | 6 | 2 | 2 | 2 | 2 | 2 |
| ASCb | A | 0 | 1 | 1 | 2 | 2 | 3 | 3 | 3 |
| MyoD | A | 1 | 4 | 1 | 7 | 4 | 4 | 4 | 4 |
| E12/E47 | A | 1 | 1 | 4 | 2 | 5 | 5 | 4 | 4 |
| Ngn | A | 1 | 1 | 3 | 3 | 2 | 2 | 3 | 3 |
| NeuroD | A | 0 | 1 | 1 | 2 | 3 | 5 | 4 | 4 |
| Atonal | A | 3 | 1 | 2 | 3 | 3 | 4 | 2 | 2 |
| Mist | A | 1 | 1 | 1 | *nf* | 1 | 1 | 1 | 1 |
| Beta3 | A | 1 | 1 | 2 | *nf* | 2 | 3 | 2 | 2 |
| Oligo | A | 0 | 2 | 3 | *nf* | 2 | 4 | 3 | 3 |
| Net | A | 1 | 1 | 2 | *nf* | 1 | 1 | 1 | 1 |
| Delilah | A | 1 | 1 | 0 | *nf* | 0 | 0 | 0 | 0 |
| Mesp | A | 1 | 1 | 0 | 7 | 4 | 5 | 3 | 3 |
| Twist | A | 1 | 1 | 2 | 2 | 4 | 3 | 2 | 2 |
| Paraxis | A | 1 | 2 | 1 | 2 | 3 | 4 | 2 | 2 |
| MyoRa | A | 1 | 4 | 1 | 1 | 2 | 2 | 2 | 2 |
| MyoRb | A | 0 | 1 | 1 | *nf* | 1 | 2 | 2 | 2 |
| Hand | A | 1 | 1 | 1 | 3 | 2 | 1 | 2 | 2 |
| PTFa | A | 1 | 1 | 1 | *nf* | 1 | 1 | 1 | 1 |
| PTFb | A | 2 | 3 | 1 | *nf* | 1 | 2 | 1 | 1 |
| SCL | A | 1 | 1 | 5 | 1 | 2 | 3 | 3 | 3 |
| NSCL | A | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| SRC | B | 1 | 1 | 0 | 3 | 3 | 3 | 3 | 3 |
| Figα | B | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| Myc | B | 1 | 1 | 1 | 5 | 3 | 6 | 4 | 4 |
| Mad | B | 0 | 1 | 1 | 5 | 3 | 4 | 4 | 4 |
| Mnt | B | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 |
| Max | B | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 |
| USF | B | 1 | 1 | 2 | 3 | 1 | 2 | 2 | 2 |
| MITF | B | 1 | 1 | 1 | 1 | 3 | 5 | 4 | 4 |
| SREBP | B | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 |
| AP4 | B | 1 | 1 | 1 | *nf* | 0 | 1 | 1 | 1 |
| MLX | B | 1 | 1 | 7 | *nf* | 3 | 1 | 2 | 2 |
| TF4 | B | 1 | 0 | 1 | *nf* | 1 | 1 | 1 | 1 |
| Clock | C | 3 | 1 | 2 | 1 | 3 | 3 | 2 | 2 |
| ARNT | C | 1 | 1 | 0 | 3 | 2 | 2 | 2 | 2 |
| Bmal | C | 1 | 1 | 0 | 2 | 2 | 2 | 2 | 2 |
| AHR | C | 2 | 1 | 1 | 2 | 3 | 4 | 2 | 2 |
| Sim | C | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 |
| Trh | C | 1 | 1 | 0 | *nf* | 1 | 2 | 1 | 1 |
| HIF | C | 1 | 1 | 1 | 4 | 2 | 6 | 4 | 4 |
| Emc | D | 1 | 1 | 2 | 5 | 4 | 5 | 4 | 4 |
| Hey | E | 1 | 1 | 1 | 1 | 2 | 4 | 4 | 4 |
| H/E(spl) | E | 11 | 11 | 12 | 15 | 6 | 15 | 8 | 8 |
| Coe | F | 1 | 1 | 1 | 3 | 3 | 5 | 4 | 4 |
| Orphan | ? | 0 | 6 | 4 | *nf* | 4 | 2 | 4 | 4 |
| Total | | 59 | 78 | 82 | 98 | 104 | 139 | 114 | 114 |

注释：*nf* 表示某基因家族在有关的研究报道中没有被发现。表中各物种 bHLH 基因数据来自参考文献[11,13-18]，其中基因家族排列的先后顺序参照了 Ledent et al. (2002)[11]。




注释：Hes 基因家族的最大似然估计进化树序列分别来自于人、小鼠、大鼠、斑马鱼、鸡和个非洲爪蟾，斑马鱼的 HEYL 作为外群(Out-group)。图中各分支上的数字为最大似然估计所得自展值(Bootstrap values)。此进化树表明，Hes 家族成员的 Hes1、Hes2、Hes3、Hes5、Hes6 和 Hes7 基因均有各自的进化起源和祖先基因。

**Figure 3. Phylogenetic tree of the H/E(spl) family**
**图 3. 脊椎动物 Hes 基因家族的系统进化分析**

是单基因或寡基因家族。45 个基因家族中，仅 10 个家族在斑马鱼、鸡、小鼠和大鼠中是单基因的家族，而文昌鱼(Lancelet)和大蛤蜊(Giant owl limpet)分别有 33 个和 24 个单基因的家族。此外，Delilah 家族在脊椎动物和大蛤蜊“丢失”，却在果蝇(Drosophila)和文昌鱼中存在。这些现象可应用分子进化中基因的生死(Birth-And-Death)理论来解释[27]。

其中，各个物种中都显示 H/E(spl)或 Hes 基因家族的成员比较多，这引起了我们的兴趣。在无脊椎动物中，该家族具有 11～12 个成员，而在已知的脊椎动物中有 6～15 个成员(表 2)。我们利用最大似然估计的方法构建人、小鼠、大鼠、斑马鱼、鸡和西方蟾蜍的 Hes 家族基因序列的进化树(图 3)，以斑马鱼的 HEYL 作为外群(Out-group)。结果发现，除人的 Hes4 外，Hes 家族成员的 Hes1、Hes2、Hes3、Hes5、Hes6 和 Hes7 基因均有各自独立的进化起源和祖先基因，这很好的验证了 Nei et al. (1997)等的基因进化的“Birth-And-Death”假说[27]。

## 4. 结论

本研究从非洲爪蟾的基因组数据库搜索鉴定出 98 个 bHLH 爪蟾转录因子。其中，有 16 个预测的转录因子(Hypothetical Protein)在数据库中注释信息不详细，在本研究中被重新发现和进一步注释，另有 2 个错误命名的基因被重新注释和命名。这些未曾研究清楚的新转录因子基因需要做进一步的分子生物学实验以深入研究结构与功能。此外，基因本体论(GO)的功能注释信息分析显示了 42 个统计显著富集分布的 GO 功能注释语句，以及各个高阶大组组内特异的显著富集分布的 GO 功能注释语句，这些注释语句均为我们认识和了解模式生物非洲爪蟾及蛙蟾类动物 bHLH 转录因子的功能、分类和基因的调控网络等生物医学研究提供了极其有用的信息。

## 5. 致谢

本文受到安徽省教育厅自然科学基金(2006KJ224B)、安徽高校优秀青年教师基金(2006jql222)和阜阳师院自然科学基金(2005QL11)联合资助，作者在这里表示感谢。

## 参考文献 (References)

[1] T. J. Boggon, W. S. Shan, S. Santagata, et al. Implication of tubby proteins as transcription factors by structure-based functional analysis. Science, 1999, 286(5447): 2119-2125.

[2] N. M. Luscombe, S. E. Austin, H. M. Berman, et al. An overview of the structures of protein—DNA complexes. Genome Biol, 2000, 1(1): 1-37.

[3] J. L. Riechmann, J. Heard, G. Martin, et al. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science, 2000, 290(5499): 2105-2110.

[4] W. R. Atchley, W. M. Fitch. A natural classification of the basic helix-loop-helix class of transcription factors. Proceedings of the National Academy of Sciences of the USA, 1997, 21(7): 5172-5176.

[5] M. E. Massari, C. Murre. Helix-loop-helix proteins: Regulators of transcription in eucaryotic organisms. Molecular and Cellular Biology, 2000, 20(2): 429-440.

[6] V. Ledent, M. Vervoort. The basic helix-loop-helix protein family: Comparative genomics and phylogenetic analysis. Genome Res., 2001, 11(5): 754-770.

[7] J. D. Stevens, E. H. Roalson, and M. K. Skinner. Phylogenetic and expression analysis of the basic helix-loop-helix transcription factor gene family: Genomic approach to cellular differentiation. Differentiation, 2008, 76(9): 1006-1022.

[8] L. Carretero-Paulet, A. Galstyan, I. Roig-Villanova, et al. Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in arabidopsis, poplar, rice, moss, and algae. Plant Physiology, 2010, 153(3): 1398–1412.

[9] C. Murre, C. P. Mc, and D. Baltimore. A new DNA binding and dimerizing motif in immunoglobulin enhancer binding, daughterless, MyoD, and Myc proteins. Cell, 1989, 56(5): 777-783.

[10] W. R. Atchley, W. Terhalle, and A. Dress. Positional dependence, cliques, and predictive motifs in the bHLH protein domain. Journal of Molecular Evolution, 1999, 48(5): 501-516.

[11] V. Ledent, O. Paquet, and M. Vervoort. Phylogenetic analysis of the human basic helix-loop-helix proteins. Genome Biol., 2002, 3(6): 301-3018.

[12] G. Toledo-Ortiz, E. Huq, and P. H. Quail. The Arabidopsis basic/helix-loop-helix transcription factor family. Plant Cell, 2003, 15(8): 1749-1770.

[13] J. Li, Q. Liu, M. Qiu, et al. Identification and analysis of the mouse basic/helix-loop-helix transcription factor family. Biochemical and Biophysical Research Communications, 2006, 350(3): 648-656.

[14] E. Simionato, V. Ledent, G. Richards, et al. Origin and diversification of the basic helix-loop-helix gene family in metazoans: Insights from comparative genomics. BMC Evolutionary Biology, 2007, 7(1): 33.

[15] Y. Wang, K. P. Chen, Q. Yao, et al. The basic helix-loop- helix transcription factor family in Bombyx mori. Development Genes and Evolution, 2007, 217(10): 715-723.

[16] Y. Wang, K. Chen, Q. Yao, et al. Phylogenetic analysis of zebrafish basic helix-loop-helix transcription factors. Journal of Molecular Evolution, 2009, 68(6): 629-640.

[17] W. Y. Liu, C. J. Zhao. Genome-wide identification and analysis of the chicken basic helix-loop-helix factors. Comparative and Functional Genomics, 2010: Article ID 682095.

[18] X. Zheng, Y. Wang, Q. Yao, et al. A genomewide survey on basic helix-loop-helix transcription factors in rat and mouse. Mamm Genome, 2009, 20(4): 236-246.

[19] U. Hellsten, R. M. Harland, M. J. Gilchrist, et al. The genome of the Western clawed frog Xenopus tropicalis. Science, 2010, 328(5978): 633-636.

[20] J. B. Bowes, K. A. Snyder, E. Segerdell, et al. Xenbase: a Xenopus biology and genomics resource. Nucleic Acids Research, 2008, 36(1): D761-D767.

[21] J. D. Thompson, T. J. Gibson, F. Plewniak, et al. The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids

Research, 1997, 25(24): 4876-4882.
[22] G. Jr. Dennis, B. T. Sherman, D. A. Hosack, et al. DAVID: Database for annotation, visualization, and integrated discovery. Genome Biol., 2003, 4(5): P3.
[23] D. W. Huang, B. T. Sherman, and R. A. Lempicki. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protocol, 2009, 4(1): 44-57.
[24] F. Ronquist, J. P. Huelsenbeck. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 2003, 12: 1572-1574.
[25] S. Guindon, O. Gascuel. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology, 2003, 52(5): 696-704.
[26] U. Hellsten, M. K. Khokha, T. C. Grammer, et al. Accelerated gene evolution and subfunctionalization in the pseudotetraploid frog Xenopus laevis. BMC Biology, 2007, 5(1): 31.
[27] M. Nei, X. Gu, and T. Sitnikova. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proceedings of the National Academy of Sciences of the USA, 1997, 94(15): 7799-7806.