**Scholars Journal of Agriculture and Veterinary Sciences** **e-ISSN 2348–1854**
Sch J Agric Vet Sci 2015; 2(1B):67-78 **p-ISSN 2348–8883**

(An International Publisher for Academic and Scientific Resources) DOI: 10.36347/sjavs.2015.v02i01.013

# Molecular Evolution of the Allergic Rhinitis Associated Gene Family *IL-33* in Mammalian Species

**Renjie Song, Feihu Wu, Ling Zhu, Gang Liu***
Anhui University of Chinese Medicine, 103 Meishan Road, Hefei-230000, China

***Corresponding Author**
**Name:** Gang Liu
**Email**: Liugang0430@126.com

**Abstract:** Interleukin-33 (IL-33) is an IL-1-like cytokine ligand for the IL-1 receptor-related protein ST2. Allergic rhinitis (AR) is one of the most common allergic inflammatory diseases. Globally, more than 600 million persons have AR. AR is divided into categories: seasonal and peren-nial. The way genomic histories drove the evolution of the IL-33 family in mammalian species was described; IL-33 genes were searched in nine mammalian such as Bos taurus, Sus scrofa, Oryctolagus cuniculus,and the Ovis aries and then through phylogenetic analysis. Microsynteny analysis was made based on genomic segments containing the IL-33 genes. Some pairs turned out to conform with syntenic genome regions, while others corresponded to those that were inverted, expanded, or contracted after the divergence of mammalians. Besides, we dated their duplications by Ks analysis and demonstrated that all the blocks were formed after the split; we observed Ka/Ks ratios representing strong purifying selections in the four mammalian species which might have been followed by gene loss and rearrangement.
**Keywords:** Allergic rhinitis, Syntenic homology, Genetic variation, IL-33 Protein

## INTRODUCTION

The serum levels of *IL-33* are significantly increased in patients with allergic rhinitis (AR) [1]. However, very little is known about the role of *IL-33* for the development of AR. *IL-33* protein was constitutively expressed in the nucleus of nasal epithelial cells and was promptly released into nasal fluids in response to nasal exposure to ragweed pollen. In human subjects , constitutive expression of *IL-33* protein in the nasal epithelial cells of healthy control subjects and down regulated expression of *IL-33* protein in inflamed nasal epithelial cells of patients with AR. *IL-33*–stimulated mast cells and basophils contributed to the early- and late-phase AR manifestation through increasing histamine release and production of chemoattractants for eosinophils/basophils, respectively.

*IL-33*, the latest member of the *IL-1* family, is the ligand for ST2 (IL-33 receptor) [2] and shares the signaling pathway with *IL-1* and *IL-18* [3].

A member of the *IL-1* family that binds to *IL-1* receptor ST2, like *IL-1 b* and *IL-18*, *IL-33*, also labeled *IL-1F11*, following *IL-1* family nomen-clature [4], is produced in a precursor form and can be cleaved by caspase-1. The binding of *IL-33* to cells that express ST2 results in the activation of NF- kB and MAP kinases. Administration of purified *IL-33* either in vitro or in vivo leads to the production of TH2-associated cytokines. Furthermore, in mice, *IL-33* induces eosinophilia, splenomegaly, and increased levels of serum Ig—leading to profound histopathological changes in the lungs and gastrointestinal (GI) tract [5].

The *IL-33* family has been studied in some model mammalian on a phylogenetic scale, but its specific evolutionary routes remain elusive. Mammalian such as *Bos taurus*, *Sus scrofa*, *Oryctolagus cuniculus* and the *Ovis aries* constitute the most important annimal family in livestock. *Mus musculus* and *Rattus norvegicus* are model systems for mammalian biology, and Bos taurus and the Sus scrofa are important meat food mammalian worldwide. Although whole-genome analyses of mammalian has demonstrated that chromosome-scale organization (so-called macrosynteny) is largely conserved [6], these studies found no traces of preservation of the specific local gene order (microsynteny) across mammalians, except in some BAC (bacterial artificial chromosome) clones [7]. As highly conserved genes mammalian, *IL-33* should also exhibit proper conserved microsynteny, which can facilitate the transfer of knowledge among related species.

## METHODOLOGY

### Genome-wide search for highly conserved *IL-33* genes in mammalian species

The most recent versions of genome, protein, expressed sequence tag (EST), and cDNA sequences for

the following nine mammalian species were downloaded from the respective genome sequence sites: *Pan troglodytes*, the *Pan troglodytes* genome sequencing project (http://uswest.ensembl.org/Pan_troglodytes/Info/Index); *Bos taurus* (version 3.5), the *Bos taurus* genome sequencing project (http://uswest.ensembl.org/Bos_taurus/Info/Index); the *Canis lupus* ,the *Canis lupus* genome sequencing project (http://uswest.ensembl.org/Canis_familiaris/Info/Index) ;*Homo sapiens*, the *Homo sapiens* genome sequencing project (http://uswest.ensembl.org/Homo_sapiens/Info/Index); *Mus musculus*, the *Mus musculus* genome sequencing project (http://uswest.ensembl.org/Mus_musculus/Info/Index); S*us scrofa*,, the *Sus scrofa* genome sequencing project (http://uswest.ensembl.org/Sus_scrofa/Info/Index);*Rattus norvegicus*, the *Rattus norvegicus* genome sequencing project (http://uswest.ensembl.org/Rattus_norvegicus/Info/Index);*Oryctolagus cuniculus*, the *Oryctolagus cuniculus* genome sequencing project (http://uswest.ensembl.org/Oryctolagus_cuniculus/Info/Index);*Ovis aries*, the *Ovis aries* genome sequencing project (http://uswest.ensembl.org/Ovis_aries/Info/Index);

Homologs were identified by blastP (E-value < $10^{-5}$ without filtering for low-complexity sequences) [8]. The dnaTools software was used to construct a local database from the nucleotide and protein sequences. Published sequences of *Homo sapiens IL-33* gene were used to search Pfam (http://pfam.sanger.ac.uk/) [9]; integrated and exact conserved *IL-33* type domain of homeobox (PF00046) and IL-33 (Interleukin 33) sequences based on the hidden Markov model (HMM) were obtained. These domain sequences were used to identify *IL-33* genes from the local databases of mammalian species. This step was a sizable one in that we aimed to identify as many similar sequences as possible. In addition, information on the chromosome locations of these genes were obtained from the results of BLASTn searches of the genome databases using the cDNA sequences of the predicted *IL-33* genes [10]. All candidate sequences that met the standards were confirmed as actual IL-33 proteins by the Pfam HMM database and SMART [11]. Finally, all of the confirmed IL-33 proteins were aligned using Clustal W [12], and all identical sequences were checked manually to remove redundant sequences.

The molecular weight (kDa) and isoelectric point (PI) of each gene were calculated using online ExPASy tools (http://www.expasy.org/tools/) [13]. The intron distribution pattern and intron/exon boundaries of the *IL-33* genes were deduced by using Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/) [14] to compare the predicted full-length cDNA or coding sequence with their corresponding genomic sequence.

**Phylogenetic analysis of *IL-33* genes**

To reconstruct the phylogeny of the *IL-33* genes across the nine mammalian species, the amino acid sequences of the IL-33 genes were subjected to parsimony analysis. A phylogenetic tree was constructed using MEGA (version 4.0) [15] and using the neighbor-joining method with the following parameters: Poisson correction, pairwise deletion, and bootstrapping(1000 replicates;random seed).All classes and subclasses were well supported, most of the classes and subclasses defined were also supported by similar gene structure of IL-33.

**Simulation of *IL-33* gene expansion patterns**

To better understand how *IL-33* genes evolved, i.e., whether they arose from a large-scale duplication events (duplicated blocks derived from whole-genome or segmental duplication) or tandem duplication, we attempted to examine the physical locations of all members of this gene family in *Homo sapiens,Pan troglodytes, Bos taurus, Canis lupus, Mus musculus,Sus scrofa,Rattus norvegicus,Oryctolagus cuniculus,*and the *Ovis aries* [16]. To categorize expansion of the *IL-33* genes, tandem duplication was characterized as the occurrence of multiple gene family members within either the same or neighboring intergenic regions. To identify large-scale duplication events, each pair of protein-coding genes (excluding non-coding RNA genes, pseudogenes, and the like) in each genome was classified as residing within a duplicated block; moreover, there must be high similarity between their neighboring protein-coding genes at the amino acid level [17]. Initially, all *IL-33* genes in each family were selected as the original anchor points representing related *IL-33* genes from different families, and then all protein-coding sequences 100 kb upstream and downstream of each anchor point were compared by pairwise BLASTp analysis to identify duplicated genes in the independent region. Finally, we tallied the total number of protein-coding genes flanking the anchor point with the best non-self match (E-value $<10^{-10}$) with a protein-coding gene neighboring the other anchor point. When five or more such gene pairs were detected, the region were considered to have originated from a large-scale duplication event.

**Microsynteny analysis**

Microsynteny analysis across the nine mammalian species was performed based on comparisons of the specific regions containing *IL-33* genes. Levels of identity and similarity between the flanking genes of each *IL-33* gene in one species and those in the other species were determined with the BLASTp program for pairwise comparison [18]. A syntenic block was defined as the region where three or more conserved homologs (BLASTp E-value $<10^{-20}$) were located within a 100-kb region between genomes [19]. The relative syntenic

quality in a region was calculated from the sum of the total number of genes in both conserved gene regions, excluding retroelements and transposons, and collapsing tandem duplications [20].

**Ks analysis of homologous segments**

The ‘‘age’’ of the duplicated gene pairs within each duplicated block or divergence of homologous segments was estimated by calculating the number of synonymous substitutions per synonymous site (Ks) between homologous genes using DnaSP (version 5.10) [21]. A sliding window analysis of nonsynonymous substitutions per non-synonymous site (Ka)/Ks ratios was conducted with the following parameters: window size, 150 bp; step size, 9 bp.

For each pair of duplicated regions, the mean Ks values for individual homologs within flanking conserved genes were calculated. Significant outliers (strongly deviating Ks values) were eliminated by Grubbs' test, also called the extreme studentized deviate method (http://www.graphpad.com/quickcalcs/Grubbs1.cfm) [22]. The Ks distributions were based on these locally averaged Ks values. Then, these Ks values could be used as the proxy for time in dating large-scale duplication events by translating them into divergence time in million years (Mya) based on a rate of $6.1 \times 10^{-9}$ substitutions per site per year. The divergence time (T) was calculated as $T = Ks/ (2 \times 6.1 \times 10^{-9}) \times 10^{-6}$ Mya [23].

**Ethical Standards**

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional guidelines on human experimentation (please name) and with the Helsinki Declaration of 1975, as revised in 2008.

## RESULTS AND DISCUSSION

**Diverse distribution of *IL-33* families in mammalian genomes**

Using the annotated genomes of *Homo sapiens, Pan troglodytes, Bos taurus, Canis lupus, Mus musculus, Sus scrofa, Rattus norvegicus, Oryctolagus cuniculus, and the Ovis aries*, BLAST and HMM searches predicted genes in the Interleukin 33. We identified 16 genes in the nine mammalian species. Based on these findings, the physical location of individual *IL-33* genes on the chromosomes was observed. The distribution of *IL-33* genes among the chromosomes in each species was diverse (Figure 1), with a slight preference toward the ends of the chromosomes (e.g., *HmChr9, PtChr9, SsChr1, RtChr1*,and so on) and with several exceptions of tandem clustered genes (e.g., *BtChr8, ClChr11, MmChr19, OaChr2, OcChr1*). In the *Bos taurus*, *IL-33* genes were found on chromosome 8; For *Canis lupus*, *IL-33* genes were found on chromosome 11. In *Homo sapiens; IL-33* genes were distributed among chromosomes 9.; *IL-33* genes in the *Mus musculus* were located among chromosome 19, and the like. These *IL-33* genes across the nine mammalian species had an open reading frame (ORF) of about 150, 230 and 270 amino acids, mean molecular weight of 17kDa, 26kDa and 30kDa.and PI of approximately 8.6. Although the distribution of these *IL-33* genes was diverse, their genetic features and biochemical properties apparently tended toward identity.

**Phylogenies and sequence structure features of *IL-33***

Based on the nucleotide sequences and by adopting the approaches of parsimony, maximum likelihood [24], and distance methods, we constructed phylogenetic trees with similar topologies in several important respects. Bootstrapping tests were performed on these trees. The generated trees were compared and the tree best supported by those methods was used for accounting for the observations (Figure 2). Combined with DNA sequence comparison, we found that PtIL-33-1 and HmIL-33-1, PtIL-33-2 and HmIL-33-2, PtIL-33-3 and HmIL-33-3 are basically the same. SsIL-33-3, BtIL-33, OaIL-33, ClIL-33-1 and ClIL-33-2 have a high degree of similarity to HmIL-33; SsIL-33-1and SsIL-33-2 are not so similar to HmIL-33 compared with other species.

Inferences of orthology based solely on phylogenetic information may appear weak, but can be supported by data concerning genomic context. This implies that the formation patterns of *IL-33* gene products are diverse.

Interleukin 33 domains are arranged in a series. Within the *IL-33* motifs (Fig. 3), we identified motif 2 and motif 3 in two species (*Pt-IL-33-3, Hm-IL-33-3*), but motifs 1, motif 2 and motif 3 can be found in the other species. Furthermore, we referred to information of the characterized functional domains of *IL-33* of human, cow, dog, and other model animals in our analysis of the protein structures of the nine mammalian species in this study. IL-33, the DNA-binding factors involved in the transcriptional regulation of key developmental processes, not only continued to emerge in these genes (Fig. 4), but also formed coiled coils in each *IL-33* gene. It has been reported that the α-helical coiled-coil structure is central to the acquisition of high-affinity DNA binding and subsequently for transcriptional activation of the corresponding genes [25]. Other functional domains of *IL-33* genes include the nuclear localization signal (NLS) and nuclear export signal for maintaining the balance of nuclear import and export.

Based on the intron-early theory [26], an excess of phase 0 introns and symmetric exons may facilitate exon shuffling by avoiding interruptions of the ORF, and consequently can accelerate the rate of re-combinational fusion and exchange of protein domains [27]. We observed an excess of phase 0 introns and

symmetrical exons,which strongly supported the intron-early theory.

**Genomic sequence comparisons and duplication events**

After identifying and sequencing *IL-33* genes from *Homo sapiens, Pan troglodytes, Bos taurus, Canis lupus, Mus musculus, Sus scrofa, Rattus norvegicus, Oryctolagus cuniculus and the Ovis aries*, we determined the chromosomal location of each gene and compared gene orders and syntenic relationships among the mammalian genomic sequences using tBLASTx comparisons [28]. This was followed by ligature of the conserved flanking genes to generate the microsynteny map of these *IL-33* genes (Fig. 5). Based on these comparisons, probable collinearity exists widely in the *IL-33* regions across the nine mammalian species. Interlaced or concordant orders of homology among the flanking genes were exhibited in the proper manner. We employed a series of procedures in an attempt to derive a better understanding of the evolution of this increasingly complex gene family. It may have undergone many processes, including gene duplication resulting from large-scale genome events (polyploidization) that spanned the entire genome, or segmental duplications (from a few nucleotides to several thousand kilobases) and local duplications (pseudogenization, neofunctionalization, and subfunctionalization) involving one or two genes, also known as tandem duplications. Given the relatively small number of *IL-33* genes in mammalians, the speculation that large-scale duplication events may have played a major role in the evolution of mammalian *IL-33* genes tends toward indeterminacy; however, segmental duplications occur in mammalian lineages as part of a "continuous" process [28].

To verify our speculation that the *IL-33* genes of the nine mammalian species were derived from large-scale duplication events, we applied tBLASTx (E-value $<10^{-10}$) against four or more of the 10 upstream or 10 downstream protein-coding genes flanking the *IL-33* genes [29]. Consequently, they produced the best match with other flanking regions of the *IL-33* genes (Figure5). Nevertheless, a few regions anchored to the *IL-33* genes had three or fewer conserved flanking genes, such as *BtIL-33* and *OcIL-33*, and *BtIL-33* and *ClIL-33*. Interestingly, some randomly selected genomic locations may also possess one or more conserved flanking genes. Therefore, the stringent criteria for eliminating the *IL-33* gene pairs residing in duplicated blocks without higher levels of divergence was defined as no more than four conserved flanking genes, and regions containing two or three flanking genes were considered flexible sets.

The alternative hypothesis is that *IL-33* genes did not evolve through duplication events, but rather via random translocations and insertional events [17]. We considered *IL-33* genes as having originated from a duplication event if they resided within a region of conserved protein-coding genes. Further investigation of the *IL-33* families revealed that three pairs of protein-coding genes flanking *OaIL-33* and *OcIL-33* were conserved; these gene pairs may have evolved from large-scale duplication and were classed in the flexible set. *BtIL-33* and *MmIL-33* contained two pairs of conserved flanking genes and therefore thought to have been involved in large-scale duplication. In *Cajanus cajan*, two protein-coding genes flanking each of two *IL-33* pairs *(OaIL-33/OcIL-33, RnIL-33/SsIL-33*) were conserved. Unfortunately, as some regions of the *Homo sapiens* genome are published in the form of scaffolds, not all of the *IL-33* genes could be localized to the chromosomes. Moreover, the sequences of the *HmIL-33-1*,*HmIL-33-2* and *HmIL-33-3* genes were identical and they were located on the same chromosome in a cluster. *ClIL-33-1*/*ClIL-33-2* were conserved, *PtIL-33-1,PtIL-33-2 and PtIL-33-3*of the *Pan troglodytes* were conserved, and the two gene pairs might also have evolved from large-scale duplication and could be classed in the flexible set.

**Conserved microsynteny of *IL-33* genes**

Microsynteny is genomics information that can be used to predict the location of homologous genes in different species. Having 80% of close homologs in the same order and transcriptional orientation was characterized as conserved microsynteny [30]. We used this to detect the evolutionary origins and orthologous relationships within the *IL-33* genes of the nine mammalian species.*IL-33* genes across the mammalian were used as anchor genes for retrospective analysis of the molecular history of the chromosomal regions on which they resided. If both ends of a *IL-33* pair could be mapped within 20 kb in the opposite orientation, they were considered microsyntenic. When the two ends were oriented opposite one another, the region was considered collinear [31]. Otherwise, the region was considered rearranged between the two species. Based on the phylogenetic relationships of *IL-33* genes among the nine mammalian species, highly conserved microsynteny in the regions across *Homo sapiens, Pan troglodytes, Bos taurus, Canis lupus, Mus musculus, Sus scrofa, Rattus norvegicus, Oryctolagus cuniculus* and the *Ovis aries* were observed by a stepwise, gene-by-gene, reciprocal pairwise comparison of the regions residing in the *IL-33* genes. Microsyntenic *IL-33* gene pairs were mapped among the nine mammalian genomes. Moreover, the conservation of microsynteny among distinct families tended to emerge as the phylogenetic tree was classified.

In the microsynteny maps, we observed a series of several-for-one microsynteny in *BtIL-33* and *MmIL-33*, *SsIL-33* and *RnIL-33*, *PtIL-33* and *HmIL-33*, and *OcIL-33* and *RnIL-33*. However, most pairs were aligned in a discordant manner. The nine species presented a higher level of microsynteny, especially *OaIL-33* and *OcIL-33,and SsIL-33 and RnIL-33*, which exhibited

stringently reverse microsynteny in their concordantly aligned flanking regions, indicating conformity of the Oa/Oc and Ss/Rn syntenic genome regions. Others pairs were aligned discordantly. *MmIL-33* and *RnIL-33*, *HmIL-33* and *ClIL-33*, *OcIL-33 and SsIL-33,and MmIL-33 and OaIL-33* were identified as having successive same-direction microsynteny. On the contrary, *OcIL-33* and *RnIL-33*, *OaIL-33* and *SsIL-33*, and *BtIL-33* and *MmIL-33* were identified as having opposite-orientation microsynteny.

According to the microsynteny maps, there was significantly less mammalian sequence within internal duplications in the nine mammalian species, measured as within-genome synteny blocks, than in synteny blocks between each of the nine mammalian genomes. The smaller amount of internal microsynteny was presumably indicative of ancient large-scale duplications that were followed by gene loss and rearrangement. These ancient large-scale duplications aided in the establishment of the high level of microsynteny observed in the comparison of the nine mammalian relative to the expectation of no prior large-scale duplication.

Microsyntenic blocks tended to be extensive but more degraded internally in the nine mammalian genomes. To estimate the extent of conserved gene content and order, we determined the synteny quality for those genes using the syntenic intervals of *Homo sapiens*, *Pan troglodytes*, *Bos taurus*, *Canis lupus*, *Mus musculus*, *Sus scrofa*, *Rattus norvegicus*, *Oryctolagus cuniculus* and the *Ovis aries*. "Synteny quality'' is calculated as twice the number of matches divided by the total number of genes in both segments after excluding transposable elements and collapsing tandem duplications, but including the conservation of genes between species [6].

We also observed that *BtIL-33*,*MmIL-33*,*OaIL-33*,*OcIL-33*,*RnIL-33 and ClIL-33* tended to form several-for-one microsynteny with Hm*IL-33, PtIL-33 and SsIL-33* , indicating that several flanking genes of Hm*IL-33*,*PtIL-33 and SsIL-33* are likely derived from one flanking gene of *BtIL-33, MmIL-33, OaIL-33, OcIL-33, RnIL-33 and ClIL-33*.This in turn indicated that duplication events, random translocations, and even insertional events might have occurred in the evolutionary history of *BtIL-33, MmIL-33, OaIL-33, OcIL-33, RnIL-33 and ClIL-33*. The distinction of duplication events among the nine mammalian species demonstrates the sophisticated in fluences and processes of evolution in mammalian.

**Dating duplication events**

Assuming synonymous silent substitutions per site (Ks) occur at a permanent rate over time, we were able to use the conserved flanking protein-coding genes to estimate the dates of the duplication events. Each pair of proteins in the microsynteny block was aligned at the amino acid level, and then codons from gapless aligned regions were used to calculate Ks values using codeml [32]. We discarded any Ks values >2.0 because of the risk of saturation [36]. The approximate date of the duplication event was then calculated using the mean Ks and an estimated rate of silent site substitutions of $6.1 \times 10^{-9}$ substitutions/synonymous site/year. The divergence time (T) was calculated as $T = Ks/(2 \times 6.1 \times 10^{-9}) \times 10^{-6}$ Mya [23]. Table 1 lists the mean Ks values for each duplication event and the estimated date. We obtained a distribution reflecting the approximate age of the duplication by plotting the number of duplication blocks against the average Ks value of the block pairs. There was a clear peak for age distribution at 0.5 synonymous substitutions per site, and a lower peak or set of peaks at 0.3, suggesting that they can all be referred to as "recent duplications".

**Strong purification selection for *IL-33* genes**

Nearly the entire *IL-33* gene family of soybean was expanded by large-scale gene duplication, such as segmental duplication, aneuploidy, or polyploidy, leading to a punctuated, dramatic increase in the number of *IL-33* genes. To understand the evolutionary constraints acting on this gene family better, we observed the Ka/Ks ratios for 34 unambiguous pairs of *IL-33* paralogs (excluding paralogs from the flexible sets). The pairwise comparison results showed that the Ka/Ks ratios of almost all of the paralog pairs were <1 (Fig. 6), suggesting that the *IL-33* gene family had mainly undergone the effect of strong purifying selections and that the *IL-33* genes are slowly evolving at the protein level.

As overall strong purifying selection could have masked positive selection at a few individual codon sites, we performed sliding window analysis of the Ka/Ks ratios between each pair of *IL-33* paralogs, which were derived from gene duplication events. Consistent with functional constraint being dominant in these domains, the *IL-33* gene domains had generally lower Ka/Ks ratios (valleys) in comparison with the regions outside them (Fig. 7). Moreover, the homeobox domains had undergone stronger purifying selections more widely than the *IL-33* domains.

**Table 1: Estimates of the dates for the synteny blocks**

| Synteny blocks of IL-33 gene | Ks | Date (mya) |
|---|---|---|
| BtIL-33&ClIL-33 | 0.33562 | 27.5 |
| BtIL-33&HmIL-33 | 0.37981 | 31.1 |
| BtIL-33&MmIL-33 | 0.58227 | 47.7 |
| BtIL-33&OaIL-33 | 0.07743 | 6.35 |
| BtIL-33&OcIL-33 | 0.56992 | 46.7 |
| BtIL-33&PtIL-33 | 0.42536 | 34.9 |
| BtIL-33&RnIL-33 | 0.59182 | 48.5 |
| BtIL-33&SsIL-33 | 0.20355 | 16.7 |
| ClIL-33&HmIL-33 | 0.333 | 27.3 |
| ClIL-33&MmIL-33 | 0.58675 | 48.1 |
| ClIL-33&OaIL-33 | 0.37402 | 30.7 |
| ClIL-33&OcIL-33 | 0.50933 | 41.7 |
| ClIL-33&PtIL-33 | 0.33897 | 27.8 |
| ClIL-33&RnIL-33 | 0.57011 | 46.7 |
| ClIL-33&SsIL-33 | 0.28833 | 23.6 |
| HmIL-33&MmIL-33 | 0.49567 | 40.6 |
| HmIL-33&OaIL-33 | 0.35869 | 29.4 |
| HmIL-33&OcIL-33 | 0.52813 | 43.3 |
| HmIL-33&PtIL-33 | 0.00571 | 0.5 |
| HmIL-33&RnIL-33 | 0.47904 | 39.2 |
| HmIL-33&SsIL-33 | 0.32817 | 26.9 |
| MmIL-33&OaIL-33 | 0.58413 | 47.9 |
| MmIL-33&OcIL-33 | 0.61999 | 50.8 |
| MmIL-33&PtIL-33 | 0.49037 | 40.2 |
| MmIL-33&RnIL-33 | 0.29046 | 23.8 |
| MmIL-33&SsIL-33 | 0.56495 | 46.3 |
| OaIL-33&OcIL-33 | 0.57816 | 47.4 |
| OaIL-33&PtIL-33 | 0.36261 | 29.7 |
| OaIL-33&RnIL-33 | 0.57224 | 46.9 |
| OaIL-33&SsIL-33 | 0.20939 | 17.2 |
| OcIL-33&PtIL-33 | 0.52531 | 43.1 |
| OcIL-33&RnIL-33 | 0.62781 | 51.5 |
| OcIL-33&SsIL-33 | 0.5321 | 43.6 |
| PtIL-33&RnIL-33 | 0.47371 | 38.8 |
| PtIL-33&SsIL-33 | 0.33204 | 27.2 |
| RnIL-33&SsIL-33 | 0.59115 | 48.5 |

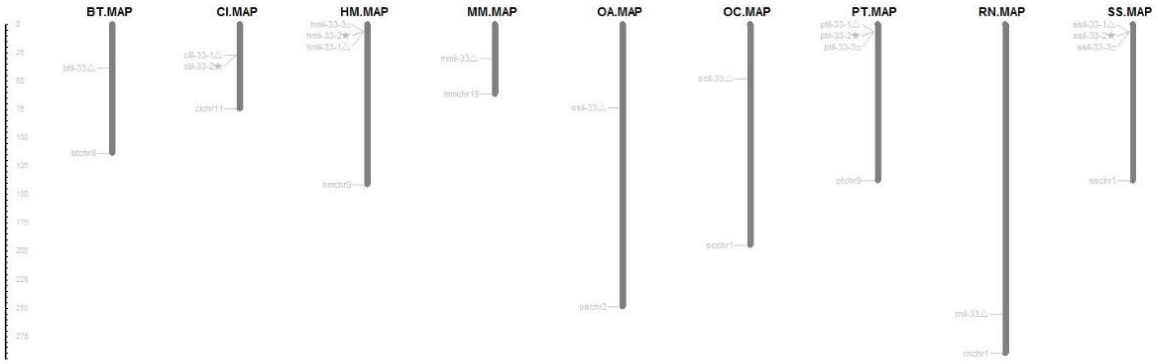


**Fig. 1: Chromosomal location of *IL-33*. The distribution of *IL-33* genes among the chromosomes in each species is diverse, with a slight tendency toward the ends of the chromosomes.**

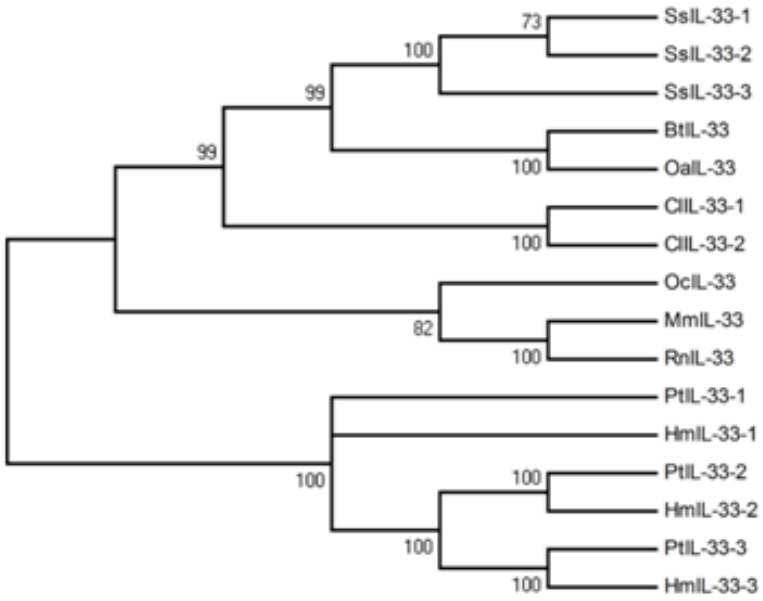


**Fig. 2: Phylogenetic tree of *IL-33* proteins from *Homo sapiens, Pan troglodytes, Bos taurus, Canis lupus, Mus musculus, Sus scrofa, Rattus norvegicus, Oryctolagus cuniculus,* and the *Ovis aries*. This tree was constructed based on amino acid sequence comparison of the conserved domain between the mammalian species by the neighbor-joining method with 1,000 bootstrap replicates. Bootstrap values are shown at each node. Different subclasses of *IL-33* are indicated.**

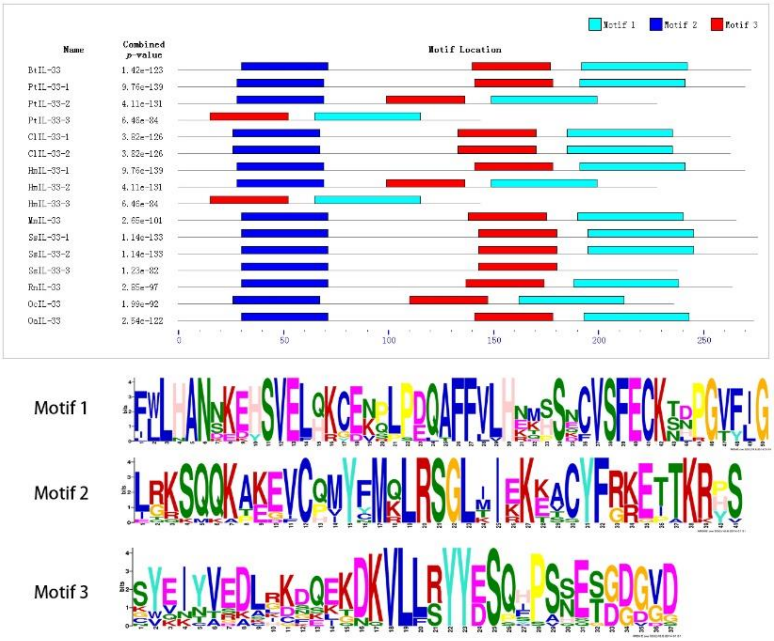


**Fig. 3: Schematic structures of mammalian IL-33 proteins. Schematic structures of 16 IL-33 proteins identified in the nine mammalian species are shown with names of all the members on the left side of the figure. Different domains are indicated by different boxes denoted. Sequence logos for IL-33 motifs in the nine mammalian species.**

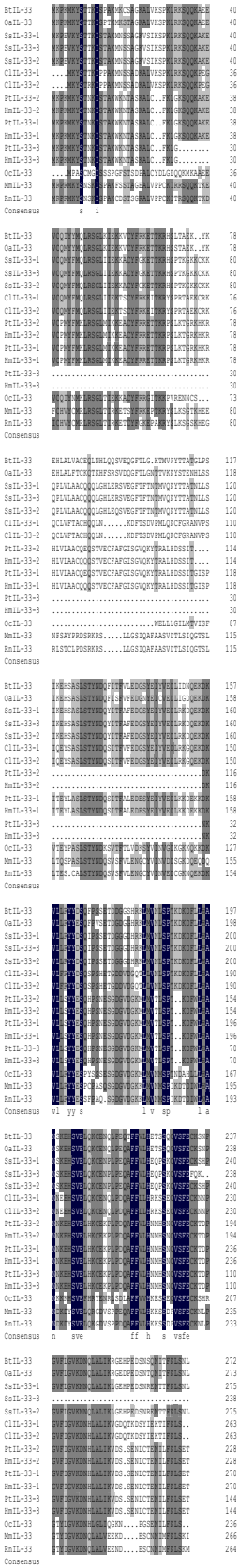

**Fig. 4: Alignment of the amino acid sequences of the *IL-33* proteins in the nine mammalian species. Dashes indicate gaps. Conserved positions in a IL consensus sequence are indicated below the alignment. Gray, conserved residues; white, residues showing variation. The three *α*-helices of the IL and leucine residues of the leucine zipper are shown on the alignment.**

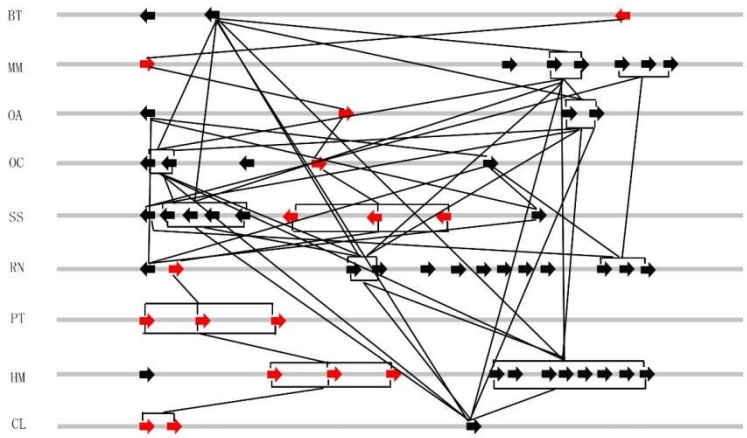


**Fig. 5: The microsynteny map of *IL-33* genes in the nine mammalian species. The relative positions of all flanking protein-coding genes are defined by the anchored *IL-33* genes (red). Gray horizontal lines, chromosome segments; arrows, individual genes and their transcriptional orientations. All genes are numbered from left to right for each segment. Duplicated genes are given the same number but ordered alphabetically. Lines connect the conserved gene pairs among the segments.**

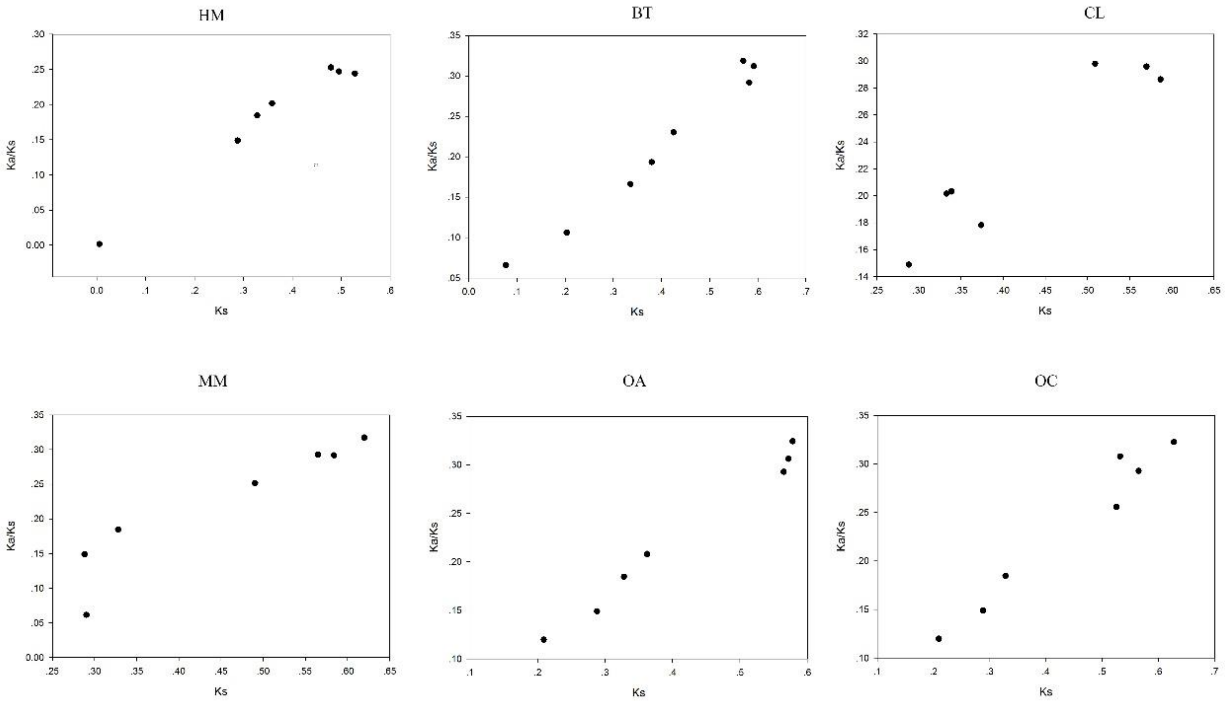


**Fig. 6: The Ka/Ks ratios of the duplicated *IL-33* genes in *HM, BT, CL, MM, OA, OC* are shown in the scatter plots. The y and x axes denote the Ka/Ks ratio and synonymous distance for each pair, respectively.**

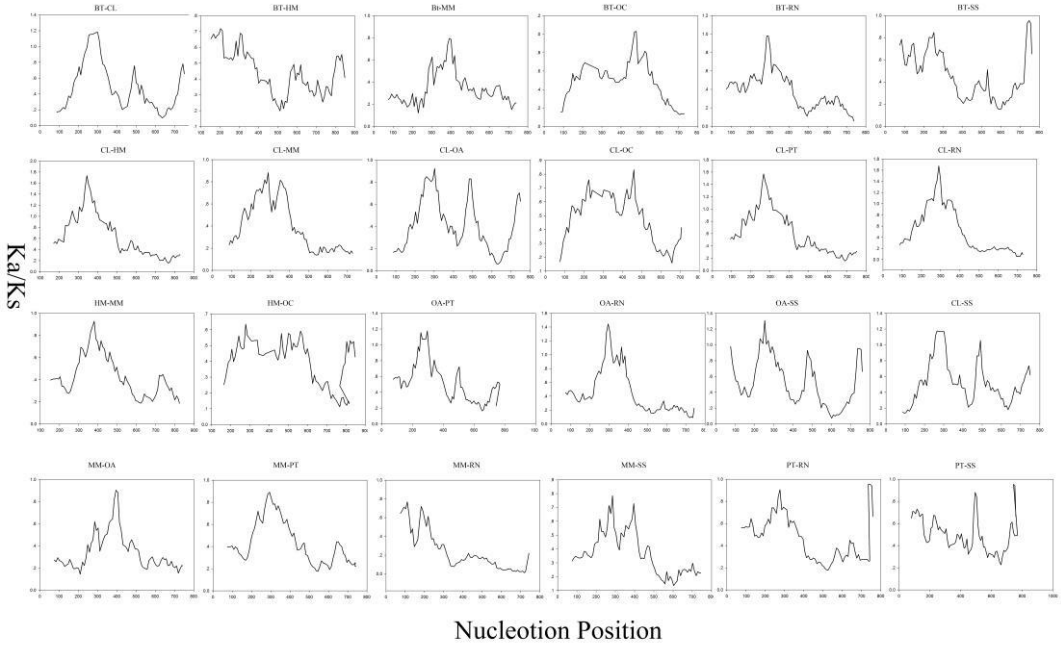


**Fig. 7: Sliding window plots of duplicated *IL-33* genes. The window size is 150 bp, and the step size is 9 bp.**

## CONCLUSION

It has been suggested that *IL-33* genes play a variety of roles in the development of AR. The gene family has expanded in mammalian with increasing organismal complexity. There were more members of the *IL-33* gene family in *Homo sapiens* (3), *Pan troglodytes* (3), *Sus scrofa* (3), *Bos taurus* (1),*Canis lupus* (1), *Mus musculus* (1), *Rattus norvegicus* (1) ,*Oryctolagus cuniculus* (1), *Ovis aries* (1). The diverse distribution among the nine mammalian species corresponds to distinct evolution processes and different surroundings, e.g., food species,biological chain,and nature disaster.

Comparisons among the *IL-33* genes across the nine mammalian genomic sequences demonstrated extensive synteny plus the existence and timing of one or more large-scale genome duplications early in mammalian evolution. There was stronger microsynteny between *Homo sapiens* and the *Sus scrofa* than between *Rattus norvegicus* and *Bos taurus*, which may be attributed to the reduced rate of evolution in *Rattus norvegicus* and *Bos taurus* since the ancient whole-genome duplication [33]. Thus, they are expected to undergo a series of evolutionary events and experience a higher rate of evolution.

Of the *IL-33* flanking gene clades, microsynteny was detected in most pairs and alignment was discordant, suggesting that these mammalian genomes have undergone extensive rearrangements, including translocations and inversions, relative to mammalian genomes. The widely present microsynteny indicated conservation across the nine mammalian species, while the opposite, discordant orientations may have been due to chromosomal rearrangement [34]. The smaller amount of internal microsynteny demonstrated that ancient large-scale duplications might have been followed by gene loss and rearrangement. The nature of internal microsynteny in the nine mammalian species provides further evidence that a large-scale duplication predated speciation.

Assuming genome duplication preceded speciation, the microsynteny maps should exhibit paired microsynteny blocks, each corresponding to the offspring of the ancient duplication event and each exhibiting comparable levels of microsynteny between the four species. In addition, if a single large-scale duplication event generated homologous segments, they should all have been created at the same time.

## ACKNOWLEDGMENTS

We extend our thanks to the reviewers for their careful reading and helpful comments on this manuscript.

## Abbreviations

Pan troglodytes (PT), Homo sapiens(HM), Bos Taurus (BT), Canis lupus (CL), Mus musculus (MM), Sus scrofa (SS), Rattus norvegicus (RN), Oryctolagus cuniculus(OC), Ovis aries(OA)

## REFERENCES

1. Sakashita M, Yoshimoto T, Hirota T, Harada M, Okubo K, Osawa Y *et al.* ; Association of serum interleukin-33 level and the interleukin-33 genetic variant with Japanese cedar pollinosis. Clinical & Experimental Allergy, 2008 ; 38(12) : 1875-1881.
2. Allakhverdi Z, Smith DE, Comeau MR, Delespesse G ; Cutting edge: The ST2 ligand IL-33 potently activates and drives maturation of human mast

cells. The Journal of Immunology, 2007 ; 179(4): 2051-2054.
3. Carriere V, Roussel L, Ortega N, Lacorre DA, Americh L, Aguilar L *et al.* ; IL-33, the IL-1-like cytokine ligand for ST2 receptor, is a chromatin-associated nuclear factor in vivo. Proceedings of the National Academy of Sciences, 2007 ; 104(1) : 282-287.
4. Sims JE, Pan Y, Smith DE, Nicklin MJ, Barton JL, Bazan JF *et al.* ; A new nomenclature for IL-1-family genes. Trends in immunology, 2001; 22(10): 536-537.
5. Schmitz J, Owyang A, Oldham E, Song Y, Murphy E, McClanahan TK *et al.* ; IL-33, an interleukin-1-like cytokine that signals via the IL-1 receptor-related protein ST2 and induces T helper type 2-associated cytokines. Immunity, 2005; 23(5): 479-490.
6. Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F, Gouzy J *et al.*; Legume genome evolution viewed through the Medicago truncatula and Lotus japonicus genomes. Proc Natl Acad Sci U S A., 2006; 103(40): 14959-14964.
7. Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J *et al.* ; Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci U S A., 2004; 101(43): 15289-15294.
8. Irimia M, Tena JJ, Alexis MS, Fernandez-Miñan A, Maeso I, Bogdanovic O *et al.* ; Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. Genome Res., 2012; 22(12): 2356-2367.
9. Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T *et al.* ; Pfam: clans, web tools and services. Nucleic Acids Res., 2006 ; 34(Database issue): D247-251.
10. Mewes HW, Frishman D, Güldener U, Mannhaupt G, Mayer K, Mokrejs M *et al.* ; MIPS: a database for genomes and protein sequences. Nucleic Acids Res., 2002 ; 30(1) : 31-34.
11. Letunic I, Doerks T, Bork P ; SMART 6: recent updates and new developments. Nucleic Acids Res., 2009 ; 37(Database issue): D229-232.
12. Thompson JD, Higgins DG, Gibson TJ ; CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res., 1994 ; 22(22): 4673-4680.
13. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD *et al.*; Protein identification and analysis tools in the ExPASy server. Methods Mol Biol., 1999 ; 112: 531-552.
14. Guo AY1, Zhu QH, Chen X, Luo JC ; GSDS: a gene structure display server. Yi Chuan, 2007 ; 29(8): 1023-1026.
15. Tamura K, Dudley J, Nei M, Kumar S ; MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution, 2007 ; 24(8) : 1596-1599.
16. Dehal P, Boore JL ; Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biology, 2005 ; 3(10) : e314.
17. Maher C, Stein L, Ware D ; Evolution of Arabidopsis microRNA families through duplication events. Genome Res., 16(4): 510-519.
18. Deleu W, González V, Monfort A, Bendahmane A, Puigdomènech P, Arús P ; Structure of two melon regions reveals high microsynteny with sequenced plant species. Mol Genet Genomics, 2007 ; 278(6): 611-622.
19. Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M *et al.*; Genome structure of the legume, Lotus japonicus. DNA Res., 2008 ; 15(4): 227-239.
20. Cannon SB1, McCombie WR, Sato S, Tabata S, Denny R, Palmer L *et al.* ; Evolution and microsynteny of the apyrase gene family in three legume genomes. Mol Genet Genomics, 2003 ; 270(4): 347-361.
21. Librado P, Rozas J ; DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 2009 ; 25(11): 1451-1452.
22. Gensel JC, Tovar CA, Bresnahan JC, Beattie MS ; Topiramate treatment is neuroprotective and reduces oligodendrocyte loss after cervical spinal cord injury. PloS one, 2012 ; 7(3) : e33519.
23. Lynch M, Conery JS ; The evolutionary fate and consequences of duplicate genes. Science, 2000 ; 290(5494): 1151-1155.
24. Guindon S, Lethiec F, Duroux P, Gascuel O ; PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res., 2005 ; 33(Web Server issue): W557-559.
25. Blazek D, Barboric M, Kohoutek J, Oven I, Peterlin BM ; Oligomerization of HEXIM1 via 7SK snRNA and coiled-coil region directs the inhibition of P-TEFb. Nucleic Acids Res., 2005; 33(22) : 7000-7010.
26. Gilbert W; The exon theory of genes. Cold Spring Harb Symp Quant Biol., 1987; 52:901–905
27. Patthy L; Intron-dependent evolution: preferred types of exons and introns. FEBS Letters, 1987; 214(1): 1-7.
28. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W *et al.*; Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res., 1997; 25(17): 3389-3402.
29. Giorda R, Bonaglia MC, Beri S, Fichera M, Novara F, Magini P ; Complex segmental duplications mediate a recurrent dup(X)(p11.22-p11.23) associated with mental retardation, speech delay, and EEG anomalies in males and females. Am J Hum Genet., 2009 ; 85(3): 394-400.
30. Vision TJ, Brown DG, Tanksley SD ; The origins of genomic duplications in Arabidopsis. Science, 2000; 290(5499): 2114-2117.
31. Mural RJ, Adams MD, Myers EW, Smith HO,

Miklos GL, Wides R *et al*.; A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. Science, 2002 ; 296(5573):1661-1671.
32. Kim H, Hurwitz B, Yu Y, Collura K, Gill N, SanMiguel P *et al*.; Construction, alignment and analysis of twelve framework physical maps that represent the ten genome types of the genus Oryza. Genome Biol., 2008; 9(2): R45.
33. Yang Z ; PAML: a program package for phylogenetic analysis by maximum likelihood. Computer applications in the biosciences: CABIOS, 1997 ; 13(5): 555-556.
34. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U *et al*.; The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science, 2006; 313(5793): 1596-1604.
35. Phan HT, Ellwood SR, Hane JK, Ford R, Materne M, Oliver RP; Extensive macrosynteny between Medicago truncatula and Lens culinaris ssp culinaris. Theor Appl Genet., 2007 ; 114(3): 549-558.
36. Blanc G, Wolfe KH; Widespread paleopolyploidy in model plant species inferredfrom age distributions of duplicate genes. Plant Cell, 2004; 16(7): 1667–1678.