The potential role of gene duplications in the evolution of imprinting mechanisms

Jörn Walter and Martina Paulsen*

Universität des Saarlandes, FR 8.2 Genetik, Postfach 151150, D-66041 Saarbrücken, Germany

Received June 30, 2003; Accepted August 14, 2003

Using the completed genomic sequences of mouse and human we performed a comparative analyses of imprinted genes and gene clusters. For many imprinted genes we could detect imprinted as well as non-imprinted paralogues. The inter- and intrachromosomal similarities between paralogues and their linkage to imprinting clusters suggests that imprinted genes were dispersed throughout the genome by gene duplications as well as translocation and transposition events. Our findings indicate that imprinting clusters may have been linked together on one (or a few) ancestral pre-imprinted chromosome(s), arguing for a common mechanistic origin of imprinting control. Imprinting may originally have evolved on a simple basis of dosage compensation required for some duplicated genes (chromosomes) followed by selection of sex-biased expression control.

INTRODUCTION

Imprinted genes are preferentially expressed from only one allele, depending on its parental origin. Silencing of one of the parental alleles is conferred by chromatin modifications, such as histone acetylation and methylation, and DNA methylation. Imprinted genes have been identified in mammalian species and higher plants (1,2) (http://cancer.otago.ac.nz/IGC/Web/home.html, www.mgu.har.mrc.ac.uk/imprinting/imprinting.html). For other clades only marginal if any imprinting effects have been described (http://cancer.otago.ac.nz/IGC/Web/home.html). Several hypothesis on the evolution of imprinting are discussed (3–5). Most of them focus on genomic imprinting in mammals, thereby implying independent evolution of imprinting in mammals and plants.

The evolutionary force that propagates genomic imprinting in mammals might be a parental conflict for the maternal resources, promoting imprinted expression of genes that function in growth regulation: paternally expressed genes may promote growth of the offspring, whereas maternally expressed genes inhibit growth for the benefit of subsequent pregnancies (4). Recently this hypothesis has been extended to behaviour of the offspring. Paternally expressed genes might promote maternal care of the daughters for the grandchildren, whereas maternally expressed genes might inhibit inbreeding (6,7). The parental conflict hypothesis is based on direct contact between embryo and mother throughout embryonic development (8). Thus, egg laying birds and monotremes should not exhibit imprinting effects. That this might indeed be the case has been shown by the finding that the orthologues of the imprinted \textit{Igf2} and \textit{Igf2r} genes are not imprinted in chicken and marsupials (9,10).

Only a few hypotheses deal with mechanistic aspects and with the molecular basis for imprinting. It has been suggested that repeated sequences and retrotransposons might be marked and silenced by allele-specific epigenetic modifications, thereby host defence mechanisms may have provided the molecular basis for genomic imprinting (3). Others propose that imprinting evolved from incomplete gene silencing, thereby leading first to random monoallelic gene expression and later to imprinting (5). Finally, imprinting and X chromosome inactivation may have co-evolved: imprinting regulatory elements may originate from duplication and translocation events of X chromosome inactivation control elements (11).

Although it is known that many imprinted genes are arranged as clusters in the mammalian genome, their physical organization has not yet been analysed in view on the evolution of imprinted gene regulation. However, this might shed some light on evolution of imprinting since physical organization is important for the interaction of genes with imprinting centres (ICs) that regulate allele-specific expression of neighbouring genes. Therefore, the physical organization of an imprinting domain should have evolved before or at the same time as imprinted gene expression.

*To whom correspondence should be addressed. Tel: +49 6813022795; Fax: +49 6813022703; Email: m.paulsen@mx.uni-saarland.de
ORGANIZATION OF IMPRINTED GENES AND THEIR PARALOGUES ON MOUSE CHROMOSOME 7

From published data it is known that about 70 imprinted genes have been identified in human and mouse. These are not evenly distributed throughout the genome but are clustered on a few chromosomes. In mouse, at least 22 imprinted protein-encoding genes reside on chromosome 7. There they are clustered in three different regions: (1) at the proximal end, containing the Peg3, Usp29, Zim1 and Zim3 genes (12); (2) the orthologue to the human Prader–Willi/Angelman syndromes domain (13,14); and (3) at the distal end the orthologue of the Beckwith–Wiedemann syndrome region (15–17). Looking in more detail on the organization of imprinted genes on mouse chromosome 7 and the orthologous regions on human chromosome 11 and 19 it becomes evident that the imprinting regions may have been linked by groups of paralogues for genes in the BWS region (Fig. 1). The Tnni2, Tnnt3 and Stk29 genes close to H19 possess paralogues that are in vicinity to an imprinted gene, i.e. Peg3 on proximal chromosome 7. In addition these paralogues, Tnt1, Tnnt3 and Kiaa1811 are flanked by the Pir-a and Pir-b genes that show also imprinting effects (18). This also involves Trpm4 on mouse chromosome 7/human chromosome 19, which is paralogous to Trpm5 in the BWS region, and also the human imprinted ZNF215 gene and its neighbour ZNF214 (19) that are located between the BWS region and ASCL3 (a Ascl2 paralogue) on the short arm of chromosome 11. The ZNF214/ZNF215 genes are in close vicinity of a cluster of olfactory receptor genes in human (20). Although the orthologous region in mouse appears to be similarly organized, murine orthologues for ZNF214 and ZNF215 appear to be absent. The vicinity of paralogues of imprinted genes to other imprinted genes and the organization of the imprinted genes on mouse chromosome 7 suggest that clustered organization and duplication events may have been involved in the molecular evolution of imprinting. Imprinting may have evolved in one genomic region (e.g. on one chromosome) and was later transmitted onto other genomic locations by duplication and translocation events.

PARALOGUES OF SOME IMPRINTED GENES ARE IMPRINTED OR LINKED TO IMPRINTED GENES

If large-scale duplication events were involved in the evolution of imprinting, it is likely that the paralogues of imprinted genes are in close vicinity to other imprinted regions. Screening the mouse genome for more paralogues of imprinted genes it becomes evident that some are also imprinted or located in close vicinity of imprinted genes in less than 4 Mb distance (Fig. 2). For the murine Ins1 that is paralogous to Ins2 in the BWS region, imprinting effects that are restricted to yolk sac have been suggested (21). Nap115 is a recently identified imprinted gene that is paralogous to Nap114 close to the BWS region (22). Nap115 resides in an intron of Herc3, a paralogue of the imprinted Ube3a gene in the PWS/AS region. However, Herc3 does not appear to be imprinted. Expanding the search for paralogues to the genes flanking the BWS region (M. Paulsen, manuscript in preparation), we identified Obpl1a, a paralogue of Obpl5, immediately adjacent to the imprinted Impact gene on chromosome 18 and Mas1, a gene neighbouring Igf2r on chromosome 17, as paralogue of the Mrge and Mrgg genes in neighbourhood of the BWS region.
Although *Mas1* is generally believed not to be imprinted, it is overlapped by a paternally expressed antisense transcript *Air* of *Igf2r* (25). In the vicinity of the imprinted genes on chromosome 17 resides also *Dll1*, a parologue of the imprinted *Dllk1* gene on chromosome 12.

A number of paralogues reside on chromosome 2 centromeric to the imprinted *Nnat* (26) and *Gnas* (27) genes: *Plagl2*, a parologue of the imprinted *Plagl1* (28), is located close to *Nnat*, *Dll4*, *Usp8*, *Trpm7* and *Snrpb* are paralogous to the *Dlk1*, *Usp29*, *Trpm5* and *Snrpn* genes and are located further
centromeric in an 11 Mb region that encompasses also the imprinted \textit{Gatm} gene (29). Although \textit{Dlk1}, \textit{Usp29}, \textit{Trpm5} and \textit{Snrpn} belong to different imprinted regions, clustering of their paralogues suggests that their ancestors were organized in close vicinity to each other. In summary our findings suggest that imprinted genes were originally linked on few chromosomal regions, and might have been dispersed in the mammalian genomes through recombination events.

\section*{RETROTRANSPOSED PARALOGUES}

A number of here described paralogues show features of retrotransposed genes, such as lack of introns, for example the \textit{Nap1l5} and the \textit{Mrgg} and \textit{Mrge} genes at the flank of the BWS region. In recent publications it has been reported that the paralogous retrotransposon-like genes \textit{Peg10} and \textit{Rtl} (\textit{Peg11}) are imprinted as well (30–32). Both genes encode two proteins with similarities to the gag and pol proteins of vertebrate retroviruses. \textit{Peg10} is located on chromosome 6 in an intron of the paternally expressed \textit{Sgc} gene. \textit{Rtl} resides in an imprinting domain on chromosome 12, encompassing the paternally expressed protein-encoding \textit{Dlk1} and \textit{Dio3} genes, and a number of maternally expressed spliced untranslated RNAs, i.e. the \textit{Git2}, \textit{Rian} and \textit{Mirtg} genes (32–34). The moderate level of similarity shared by \textit{Peg10} and \textit{Rtl} suggests that they derived from a rather ancient duplication event.

In a recent publication (35) paternal expression of the \textit{Mknn1-p1} gene has been described. This gene is a pseudogene on mouse chromosome 5 derived from the biallelically expressed \textit{Mknn1} gene. Also the paternally expressed \textit{Mrknn3} (\textit{Znf127}) in the PWS/AS region appears to be derived from the \textit{Mknn1} gene by a retrotransposition event that predate the divergence of human and rodent lineages (36).

\textit{Peg10/Rtl, Nap1l5/Nap1l4, Mknn3/Mknn1} and \textit{Masl/Mrgg/Mrge} are conserved in human and mouse indicating retrotransposition events before radiation of the human and rodent lineages. These identified paralogues are either imprinted, or in vicinity to imprinting domains, suggesting that regulatory elements important for imprinted gene expression can be transmitted by retrotransposition events. Clearly retrotransposed elements play an important role in the clustering and chromosomal organization of imprinting domains. The organization of some retrotransposed imprinted genes suggests that they were integrated into an imprinted gene cluster before duplications of such clusters. The dispersion and rearrangement of imprinting genes and clusters may subsequently have been triggered by recombination at such retrotransposed highly similar genes (elements).

Moreover transposition and recombination at retrotansposable elements may have also triggered the evolution of diversity of imprinting domains among mammalian species. There is no evidence for an orthologue of the \textit{Mbkn1-p1} pseudogene in homologous position in the human genome. The DNA sequence of the \textit{Mbkn1-p1} pseudogene is also highly similar to the \textit{Mbkn1} gene indicating a recent duplication event. Interestingly, the retrotransposed \textit{Mbkn1-p1} and \textit{Mrknn3} are imprinted, whereas the source of both genes, \textit{Mbkn1} is not. Similarly, the murine imprinted \textit{U2af1-rs1} gene appears to originate from a rodent-specific retrotransposition event (37).

There is also no indication for imprinting of the source gene. This implies that retrotransposition events might cause de novo imprinted expression of the retrotransposed gene and that the silencing mechanism is able to distinguish the endogenous gene and the retrotransposon.

Interestingly, all described imprinted retrotransposon like genes are paternally expressed, indicating that an anti-silencing mechanism specific for the male germ-line might be involved or that the genes are silenced in both germ-lines, but only the paternal alleles are reactivated by the paternal demethylation in the zygote shortly after fertilization. This would be in line with the proposal that maternal imprints at imprinting domains are more stable than paternal ones (38). It would be interesting to see whether mammals which supposedly do not show the parental asymmetry of zygotic demethylation, like sheep and canines, do possess imprinted retrotransposed genes.

\section*{DUPLICATIONS MIGHT INDUCE IMPRINTING EFFECTS}

In summary our findings suggest that duplication and translocation events play an important role in the evolution and possibly also the origin of imprinted domains. Several model hypothesis are in line with these observation. First, duplications of imprinting domains might have transmitted regulatory elements necessary for imprinting onto different genomic regions (Fig. 3). Imprinted gene regulation would have existed before duplication, and was simply transmitted onto both duplicated domains (Fig. 3A). Later chromosomal rearrangements could have caused the dispersion of clusters and the formation of different imprinting domains. Thereby previously monoallelically expressed genes might have been biallelically activated, whereas biallelically expressed genes might have been silenced on one parental allele. Our findings suggest that the dispersed imprinted gene clusters found in mammalian species might have been linked together on an ancestral imprinted chromosome. Imprinting on such an imprinted autosome might have operated similar to X-chromosome inactivation. However, in the course of dispersion, different local silencing elements were evolved and transmitted by translocation/duplication events into new environments thereby leading to differences in mechanisms of cluster control as seen today for different imprinting domains.

In a second model, imprinting originates from random monoallelic expression of paralogues (Fig. 3B): whereas genome-wide duplication leading to a tetraploid organism does not change the balance in overall gene dosage, regional duplications produce imbalanced changes in gene dosage. Therefore silencing of additional gene copies (dosage compensation) might be often crucial for survival of the affected organism. The silencing might be mediated by epigenetic modifications and might involve pairing events among the duplicated genes, thereby inducing processes such as de novo methylation (39). If this affected all duplicates of a gene its complete silencing might be also lethal for the organism. Therefore, a counting mechanism might have resulted in random monoallelic silencing of the paralogues. Subsequently, for many duplicated genes divergence of paralogues might have led to different functions and differences in tissue-specific gene expression,
thereby making monoallelic expression obsolete. Thus, for many genes the second allele might have been reactivated. However, in mammals the functional advantages of imprinted gene expression might have resulted for some genes in a change from random silencing of one allele into germ-line-specific silencing, as suggested by the parental conflict hypothesis. Many of the duplication events described are also visible in the fugu genome, and therefore predate the evolution of mammals (40) (M. Paulsen, manuscript in preparation). Hence, random monoallelic expression of paralogous genes might have persisted for a long time in early vertebrates, implying that these genes might be still monoallelically expressed in other non-mammalian clades as indicated by the observed monoallelic Igfl2 expression in chicken (41). Imprinting in mammals might have been supported by addition of a few more factors to the necessarily complex machinery of epigenetic gene regulation, for example the evolution of non-coding RNA genes such as H19 and Gtl2 and of antisense transcripts of imprinted genes as well as most probably germ line-specific marking machineries such as DNMT3L.

On the basis of such ancestral dosage-compensation/imprinting mechanism the generation of retrotransposon-derived imprinting effects most probably occurred later in the mammalian lineage, promoting the ongoing battle of sex-specific gene control. Hence, evolution of imprinting is an on-going process and may therefore have also some input in mammalian speciation.

ACKNOWLEDGEMENTS

We thank D. Prawitt for helpful discussion and S. Tierling and S. Zinßer for critical reading of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (WA 1029/3-1).

REFERENCES


Figure 3. Involvement of duplication events in evolution of imprinting. Maternally expressed genes are shown in red, paternally expressed genes in blue, biallelically expressed genes in white, monoallelically expressed genes in blue/red. (A) Duplication of an existing imprinting domain. Chromosomal rearrangements after duplication lead to a changed structure of imprinted regions. (B) Duplication of biallelically expressed genes leads to random monoallelic expression of the resulting paralogues. During evolutionary divergence of the paralogues some genes get biallelically activated, for others random monoallelic expression is shifted to imprinted gene expression.