UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES
Section de Biologie
Département de Génétique et Evolution
Professeur Denis DUBOULE

Evolution of *Hoxd* gene regulation

THÈSE

présentée aux Facultés de médecine et des sciences de l'Université de Genève
pour obtenir le grade de Docteur ès sciences en sciences de la vie,
mention Biosciences moléculaires

par
Aurélie HINTERMANN

de
Beinwil am See (AG)

Thèse N° 117

GENÈVE
Centre d'impression de l'Unige
2021

LIFE SCIENCES PhD SCHOOL GENEVA

DOCTORAT ÈS SCIENCES EN SCIENCES DE LA VIE DES
FACULTÉS DE MÉDECINE ET DES SCIENCES
MENTION BIOSCIENCES MOLÉCULAIRES

Thèse de Madame Aurélie HINTERMANN

intitulée :

«EVOLUTION OF HOXD GENE REGULATION»

Les Facultés de médecine et des sciences, sur le préavis de Monsieur D. DUBOULE, professeur ordinaire et directeur de thèse (Département de génétique et évolution), Monsieur G. ANDREY, professeur assistant (Département de médecine génétique et développement, CMU), Monsieur D. LUPIANEZ, docteur (Max Delbrück Center for Molecular Medicine in the Helmholtz Association Berlin-Mitte, Berlin, Germany), autorisent l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 7 septembre 2021

Thèse - 117 -

Le Doyen
Faculté de médecine

Le Doyen
Faculté des sciences

N.B. - La thèse doit porter la déclaration précédente et remplir les conditions énumérées dans les "Informations relatives aux thèses de doctorat à l'Université de Genève".

Table of contents

ACKNOWLEDGEMENTS 8

SUMMARY 12

RESUME 14

INTRODUCTION 16

EARLY ANIMAL EVOLUTION AND DEVELOPMENTAL GENOMICS 16

The evolution of Metazoan 16

Developmental genes 18

EVOLUTION OF LONG-RANGE REGULATION IN VERTEBRATES 19

The origin of vertebrates 19

Genes remain, regulations evolve 20

Regulatory landscapes and synteny 21

Chromatin architecture 22

***Hox* GENES TO STUDY REGULATORY MECHANISMS AND INNOVATIONS 24**

Hox genes in flies and in vertebrates – conservation beyond sequence 25

Hox gene evolution in Metazoa 26

Pleiotropy leads to increased cluster organization 27

The mouse *HoxD* locus 30

SCOPE OF THE THESIS 31

SECTION I 33

EVOLUTION OF GENOMIC ORGANIZATION AT VERTEBRATE *HOX* CLUSTERS 33

INTRODUCTION 33

RESULTS 38

DISCUSSION 43

MATERIAL AND METHOD 46

SECTION II 51

DEVELOPMENTAL AND EVOLUTIONARY COMPARATIVE ANALYSIS OF A *HOXD* REGULATORY LANDSCAPE IN MAMMALS AND BIRDS 51

ABSTRACT 54

INTRODUCTION 55
RESULTS 57
DISCUSSION 62
MATERIAL AND METHODS 65
LEGENDS TO FIGURES 69
LEGENDS TO SUPPLEMENTARY FIGURES 71
BIBLIOGRAPHY 74

CONCLUDING REMARKS 92

ANNEX I 93

BIBLIOGRAPHY 107

List of figures

Figure 1 - Phylogenetic tree of Metazoa. 17

Figure 2 - Evolution of *Hox* clusters in vertebrates. 29

Figure 3 - The genomic structure of the mouse *HoxD* locus. 34

Figure 4 - Emergence of vertebrate synapomorphies. 36

Figure 5 - *HoxD* loci in chordates. 39

Figure 6 - The four paralogous mouse *Hox* loci. 42

Hintermann et al. Figure 1 81

Hintermann et al. Figure 2 82

Hintermann et al. Figure 3 83

Hintermann et al. Figure 4 84

Hintermann et al. Figure 5 85

Hintermann et al. Figure S 1 86

Hintermann et al. Figure S 2 87

Hintermann et al. Figure S 3 88

Hintermann et al. Figure S 4 89

Hintermann et al. Figure S 5 90

Hintermann et al. Figure S 6 91

List of tables

Table 1 – List of *Hox* or *HoxD* loci in different species. 46

Table 2 – Mouse homologs in different chordate species. 47

Table 3 - genomic coordinates used to plot the different loci. 49

Table 4 – gene annotation at the four mouse *Hox* loci. 50

Table 5 – TAD sizes at the *HoxA* and the *HoxD* clusters. 50

ACKNOWLEDGEMENTS

I would like to thank Denis Duboule for letting me conduct my PhD thesis in his laboratory, for fostering such a work atmosphere of high scientific level and for providing us with both the freedom to work following our own initiatives and with many opportunities to gain experience.

I want to thank Guillaume Andrey and Darío Lupiáñez for accepting to be my thesis jury. I look forward to receiving feedbacks from such inspiring scientists.

I would like to thank Leonardo Beccari with whom I discovered Science and collaborated in most of my work and had many fascinating scientific conversations.

I want to acknowledge the fantastic support of Chase Bolt and Lucille Delisle, not only in the proofreading of this thesis manuscript, but over the years. Both of them offered their invaluable help spontaneously. Chase taught me how to determine the structures of chromatin domains and of scientific research and Lucille how to use and understand series of bioinformatics tools, together with the biology behind.

I am very thankful to Sandra Gitto, an extremely talented, motivated and hardworking research assistant who was endowed with the sense of both practice and humor.

Thanks to Bénédicte Mascrez for her great scientific advices, to Corinne Matthey for her irreplaceable help with the administrative twists and turns and to Gregory Loichot for his inputs on how to create a coherent and visible professional profile on social networks.

I would like to warmly thank Isabel Guerreiro, József Zákány, Hanh Nguyen Huynh, Fabrice Darbellay, Celia Bochaton, Rita Amândio and many more past or present members of the unige and of the EPFL labs. They all contributed to generate a great work atmosphere and to conserve critical thinking. Thanks to them together, there has not been a single day that I wished to not go to work. It sounds exaggerated, but it is not.

I appreciated the numerous scientific conversations with Laura Iglesias and with Joel Tuberosa, to open my scientific horizon.

I would like to acknowledge the inestimable contribution, even if indirect, of my family. My daughter Eva, an inexhaustible source of joy, noise and tiredness, helped me to learn how to keep a balanced life and to remain focused and efficient in my work. I want to thank my mother and her husband, my father and my sister Odile, who encouraged me and took care of Eva a countless number of times.

I thank Julien for his non-scientific inputs which helped me formulate my work in a way that I could share and explain outside of the scientific community, and for all the laughs that are essential to relax after work.

Finally, I am grateful to Sonia Vidal who guided me through the obscure paths of my own brain. She accompanied me since I started to make a better place to be of my own life, for example by belonging to a highly stimulating and fascinating scientific team.

SUMMARY

The phenotypic diversity found in vertebrates is closely related to modifications in the transcription patterns of developmental genes. Developmental genes often participate in several patterning processes and are thus very well conserved. In contrast, their associated *cis*-regulatory elements (CREs) function in a mostly time and tissue specific manner. Genetic modifications within CREs may thus alter the pattern of expression of their target developmental gene in a spatiotemporal limited manner. Consequently, they may lead to morphological changes with less probability of associated lethality than mutations affecting coding sequences. For this reason, understanding the mechanisms of CREs evolution is key to gain insight into vertebrate diversification.

In vertebrate genomes, the multiple CREs controlling the activity of common developmental target gene(s) are generally found in large genomic regions called 'regulatory landscapes'. In addition, the limits of these regulatory landscapes often correspond to the boundaries of Topologically Associating Domains (TADs), suggesting a functional link between regulatory activities and chromatin organization.

The mouse *HoxD* locus represents an excellent system to study evolution of gene regulation because it is conserved and used in multiple tissue types through development. In addition, the different regulatory mechanisms controlling *Hoxd* gene expression are physically divided by function into several regulatory landscapes, C-DOM and T-DOM; where the T-DOM is further subdivided. Remarkably, the different regulatory domains appear to match the sub-compartmentalization of the 3D genome. This correspondence provided an entrance to systematically test the relationship between genomic organization and gene regulation in an evolutionary perspective.

In this work, we started by comparing some aspects of the *Hox* genomic organization in different chordate species and at different mouse *Hox* loci. This analysis created a framework for the interpretation and contextualization of the ensuing experiments.

Then, we turned to an analysis of the transcription of anterior *Hoxd* genes in the mouse primordia of vibrissae follicles and somites, which we compared to the corresponding chicken tissues, the primordia of feather follicles and somites respectively. These structures arose independently over an enormous amount of time – approximately 300 million years separates the appearance of somites (chordates) and the evolution of the first skin appendages

(tetrapods). Yet, we found that the transcription of *Hoxd* genes is controlled by regulatory elements located in the T-DOM that are sometimes common between the two tissue types. In addition, we show that CREs regulating *Hoxd1* are systematically found in the same genomic region, which establish frequent interactions with its promoter, no matter the species (mouse, chicken) or the tissue analysed (skin primordia, somites). These findings are detailed in a first manuscript for publication, which is attached to this thesis.

Our analysis about *Hoxd1* regulation in mouse embryos led to a collaboration with the group of Aurélien Capitan (INRAE), which is attached to this thesis as an annex (Allais-Bonnet et al. 2021, Annex 1). This work showed that the presence of supernumerary horns in goats and sheep (polyceraty) was linked to alterations of *HOXD1* function. The mapping of *POLYCERATE* alleles in four-horned goats revealed that the mutation responsible for this morphological alteration is a large deletion located in the T-DOM, complementing our findings about the evolution of *Hoxd1* regulation.

Altogether, our results highlight the deep conservation of the *HoxD* cluster topology throughout vertebrates, despite divergent nucleotide sequences and regulatory activities. They support and reinforce the proposal that 3D chromatin organization both facilitates the emergence of new enhancers and limit their range of action.

RESUME

La diversité phénotypique des vertébrés est fortement liée à des variations dans les patterns de transcription des gènes du développement. Ces gènes sont généralement impliqués dans plusieurs processus d'organogenèse et ont par conséquent été fortement conservés au cours de l'évolution. En revanche, leurs séquences régulatrices, appelées *cis*-regulatory elements (CREs), ont une activité majoritairement spécifique aux tissus et aux stades de développement. Ainsi, certaines modifications de ces séquences de régulation affectent le pattern d'expression de leur gène cible de manière limitée dans le temps et dans l'espace. De ce fait, elles sont plus susceptibles de provoquer des altérations morphologiques non létales que des modifications des séquences codantes. Une compréhension accrue des mécanismes entraînant l'évolution des séquences régulatrices permettrait donc d'améliorer nos déductions quant aux raisons de l'incroyable diversité biologique des vertébrés.

Les génomes de ces animaux présentent des domaines de chromatine particuliers, appelés 'paysages de régulation'. Ces paysages rassemblent une multitude de séquences régulatrices qui contrôlent à distance l'activité d'un ou plusieurs gènes de manière coordonnée. L'étendue des paysages de régulation coïncide souvent avec celle de structures topologiques d'interactions préférentielles (TADs pour 'topological associating domains') au sein desquelles les séquences ont une probabilité accrue de se retrouver proches les unes des autres. Cette correspondance a conduit à l'idée que l'organisation fonctionnelle de la régulation génique à distance et l'organisation physique de la chromatine dans l'espace du noyau forment deux aspects interdépendants du génome.

Le locus *HoxD* de la souris constitue un système adéquat pour étudier l'évolution de la régulation génique parce qu'il est fortement conservé et parce qu'il est actif dans de nombreux processus développementaux. Par ailleurs, les différents mécanismes de régulation qui contrôlent l'expression des gènes *Hoxd* sont compartimentés en différents paysages de régulation, le C-DOM (pour domaine centromérique) et deux domaines imbriqués plus petits formant le T-DOM (pour domaine télomérique). Chacun de ces paysages présente une structure topologique en TAD, formant un ensemble génomique qui permet de tester de manière systématique la corrélation entre l'organisation de la chromatine et la régulation génique dans une perspective évolutive.

Pour ce travail, nous avons d'abord comparé certains aspects de l'organisation génomique entre des complexes *Hox* de différentes espèces de chordés, ainsi qu'entre les différents clusters *Hox* chez la souris. Cette analyse nous a permis de formuler un cadre théorique pour l'interprétation et la contextualisation des expériences suivantes.

Nous avons ensuite analysé la transcription des gènes *Hoxd* antérieurs dans les primordia de vibrisses et les somites chez la souris que nous avons comparés aux tissus correspondant chez le poulet, à savoir les primordia de plumes et les somites. Les somites (chordés) et les premiers appendices cutanés (tétrapodes) sont apparus indépendamment sur terre il y a plus de 300 millions d'années. Nous avons pourtant montré que la transcription des gènes *Hoxd* dans ces structures est contrôlée par des éléments de régulation situés dans le T-DOM, et qui sont parfois communs entre les deux tissus. Nous montrons également que la régulation à distance de *Hoxd1* dépend systématiquement d'une région génomique commune qui établit de fréquentes interactions avec son promoteur, indépendamment du type de tissus ou de l'espèce considérée. Les résultats de cette recherche sont détaillés dans un premier manuscrit pour publication, attaché à cette thèse.

L'analyse de la régulation de *Hoxd1* dans l'embryon de souris a conduit à une collaboration avec le groupe de recherche d'Aurélien Capitan (INRAE), qui est annexée à cette thèse (Allais-Bonnet et al. 2021, Annex 1). Ce travail montre que la présence de cornes surnuméraires chez les chèvres et les moutons, des animaux qualifiés de polycérates, est associée à des défaillances dans la fonction du gène *HOXD1*. La cartographie des allèles *POLYCERATE* chez la chèvre a permis de localiser la mutation responsable de cette anomalie anatomique au sein du T-DOM, ce qui complémente nos résultats quant à l'évolution de la régulation de *Hoxd1*.

Nos résultats de recherche, considérés dans leur ensemble, mettent en évidence la forte conservation du complexe *HoxD* chez les vertébrés, et ce malgré de fortes divergences dans les séquences ADN ainsi que dans les activités régulatrices. Nos résultats soutiennent et renforcent l'hypothèse que l'organisation de la chromatine au sein du noyau facilite l'émergence de nouvelles régulations, tout en limitant leur portée.

INTRODUCTION

EARLY ANIMAL EVOLUTION AND DEVELOPMENTAL GENOMICS

The evolution of Metazoan

The early history of animals is marked by the diversification and complexification of their forms, physiologies and behaviours. According to our current knowledge of metazoan evolution, Porifera (sponges) diverged first, followed by Cnidaria (jellyfishes and corals) and finally Bilateria (**Figure 1**). The "Cambrian Explosion" (~540-520 MYA) describes the epoch when the Bilateria radiated, which led to the appearance of all major groups of living animals, as reflected by the remarkable fossil record of this period (Briggs 2015). Bilateria are named after their bilaterally symmetric body plan presenting a left side that is a mirror image of the right side, which was an addition to the already existing anteroposterior and a dorsoventral body axes. Prior to this, animals were radially symmetric and produced two germ layers, the endoderm and the ectoderm. With few exceptions, Bilateria develop a third germ layer between the two pre-existing ones, the mesoderm. The mesoderm enabled the building of complex animals by providing numerous new types of tissues like muscles, cartilage or bones (Burke 2007). For example, the development of a jawed head (anatomy) in gnathostomes is tightly associated with the appearance of active predation (behaviour) and the need for dedicated digestive and excretory systems, as well as a complex nervous system (physiology) (reviewed in Martinez-Morales 2016).

Morphology endures fossilization processes and is thus generally used to try to understand the evolutionary histories of organisms, by inferring ancestral anatomies, behaviours and physiologies. Morphology is established during embryonic development, which defines the construction of a whole, complex organism from a single cell. An important event during the course of animal evolution is the emergence of segmentation, a developmental process that forms repeated units along the anteroposterior body axis. It characterizes annelid, arthropod and chordate lineages, which most probably developed segmentation independently. Except for annelids, whose different body units are a repetition of the same sets of organs, segmented phyla are among the most diverse groups.

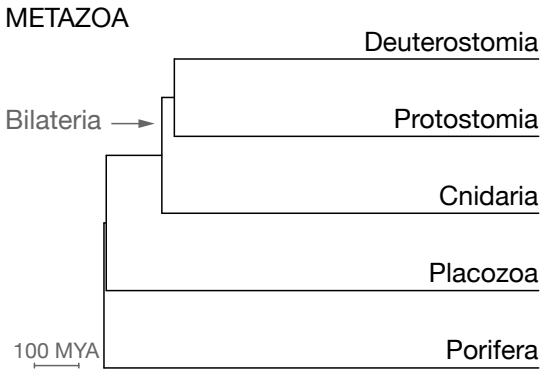


Figure 1 - Phylogenetic tree of Metazoa.

Phylogenetic tree showing major phyla of Metazoa (animals). Phylum names are indicated on the branches of the tree. According to the current metazoan phylogeny, Porifera (sponges) diverged first, followed by Placozoa (e.g., *Trichoplax adhaerens*, flat, multicellular and marine organism), then Cnidaria (jellyfishes, anemones, hydras and corals) and finally Bilateria. Bilateria are characterized by a bilateral symmetry and the presence of a third germ layer, the mesoderm. Most Bilateria have a mouth separated from the anus. The blastopore is the first opening of bilaterian embryos and becomes the mouth in Protostomia (e.g., mollusks, arthropods), while it becomes the anus in Deuterostomia. Chordata (chordates) belong to Deuterostomia, together with Hemichordata (e.g., acorn worms) and Echinodermata (e.g., sea stars).

Indeed, the modular properties of a segmented body plan are assumed to confer increased evolvability, by providing the possibility to change one part of the body without affecting the others, or the whole organism. At first relatively uniform, embryonic segments then differentiate into a variety of anatomical structures. Therefore, modifications of those segments may result in new body plans, as exemplified by the wide variations observed not only in the number but also the type of vertebrae in different species. For example, some frogs only have 6 pre-sacral vertebrae while in caecilians, a group of limbless amphibians, this number reaches nearly 300. Snakes have hundreds of thoracic vertebrae while platyfishes are mostly made of caudal vertebrae (Richardson et al. 1998).

Developmental genes

Developmental biology evaluates the processes by which an embryo is organized in a coherent ensemble throughout its development. It particularly addresses the relationships between genetics and morphology. The products of developmental genes include transcription factors, which regulate the activity of other genes, and signaling molecules, mediating information between cells. Combinations of developmental gene activities provide cells with their spatiotemporal location and hence, their functional properties. Consistently, expression domains of developmental genes often correlate with described embryonic regions or tissues, like somites, limb bud, or skin placodes, and changes in those domains may lead to developmental phenotypes (reviewed in Bolt and Duboule 2020).

Evolutionary biology generally consists of formulating and refining scenarios that could explain what forms were selected from the set of possible phenotypes generated during development. The discovery of *Hox* genes by Ed Lewis in 1978, followed by their detailed characterization in many species and developmental contexts, allowed some novel interpretations about the molecular evolution of metazoan genomes. Namely, the concept of a genetic toolkit (i.e., Duboule and Wilkins 1998; Meyerowitz 1999; Levine and Tjian 2003; Carroll 2008) synthetized the fact that the number and types of developmental genes are limited. Therefore, it was proposed that animal diversity and complexity principally result from differences in the developmental gene regulation rather than in their protein coding sequences (for a historical perspective, see King and Wilson 1975). In vertebrates, developmental genes became highly pleiotropic, probably due to major genomic reorganizations that occurred early in this lineage and paralleled by the multiplication of their regulatory inputs.

In summary, the current interpretation of evolution explains the phenotypic diversity of animals by modifications of regulatory control of otherwise largely conserved developmental genes.

EVOLUTION OF LONG-RANGE REGULATION IN VERTEBRATES

The origin of vertebrates

The earliest vertebrates appeared during the early Cambrian Period and then radiated in a variety of morphologies and sizes, from *Paedophryne* frogs measuring around 7 mm (Rittmeyer et al. 2012) to blue whales that can reach 30 m. Typical vertebrate innovations include a jawed head on the top of the trunk and a bony endoskeleton. As they populated aquatic, terrestrial and aerial environments, their anatomical diversification was accompanied by changes to other biological processes, like adaptive immune system, elaborated sensory organs or expansion of the central nervous system. Although not exact, the complexity of living forms is often approximated by the number of different cell types constituting an organism (Levine and Tjian 2003).

Considering fossil records, molecular clocks and comparative genomics with the invertebrate chordate *Amphioxus* (Marlétaz et al. 2018), it was established that the origin of vertebrates is tightly related to major changes in the genome of their common ancestor. In his seminal book “Evolution by gene duplication”, Susumu Ohno proposes that two successive whole genome duplications (WGDs) occurred between the emergence of cephalochordates and jawed vertebrates (gnathostomes) (Ohno,S 1970). He introduces the idea that genetic redundancy resulting from DNA sequence duplications offers opportunities for the emergence of new gene functions (neo-functionalization) by allowing one gene copy to escape purifying selection. This scenario, commonly called the 2R-hypothesis for 2 rounds of WGD, has been refined and debated since then (reviewed in Kasahara 2007). Ohno formulated his theory from observations of genome sizes, karyotypes and naturally occurring tetraploid fish and amphibian species. The advent of next-generation sequencing (NGS) provided important evidences supporting Ohno’s hypothesis (i.e., Dehal and Boore 2005; Simakov et al. 2020). For example, it revealed that for each invertebrate gene, up to four copies can be found in vertebrate species (Garcia-Fernàndez and Holland 1994). It also exposed the conservation of duplicated syntenic regions, which was used to unambiguously asses homology between vertebrate and *Amphioxus* genetic loci (Touceda-Suárez et al. 2020).

Genes remain, regulations evolve

Gene neo-functionalization frequently occurs *via* the establishment of new regulatory relationships, which are facilitated by genetic redundancy. Long before the advent of NGS, namely in the paper of Mary-Claire King and A. C. Wilson (King and Wilson 1975), it was already proposed that interspecies variability would result from modifications affecting the regulation of organism macromolecules. At the time, in an effort of the scientific community to measure the genetic distance between species, several molecular methods were used to compare orthologous proteins and nucleic acids. As no or very few differences were observed, it was proposed that variations in the regulation of these molecules, and not in their structure, could account for the observed biological differences, in this particular case, between chimps and humans.

Vertebrate developmental genes are often highly pleiotropic, meaning that they are involved in the formation of many structures. The multiple functions of developmental genes result from a complex transcriptional control achieved by combinations of *cis*-regulatory elements (CREs) (e.g., (Spitz and Furlong 2012; Bolt and Duboule 2020). CREs are particularly numerous in vertebrate lineages (Marlétaz et al. 2018). They include promoters, enhancers, repressors and binding sites for architectural proteins, which actions impact on the probability of transcriptional initiation. CRE activities are mostly time and tissue-specific and hence, the complete expression pattern of pleiotropic genes results from the cumulation of all their cognate regulatory influences. The complexity of a gene function can roughly be estimated by the number of different structures in which it is expressed, correlating with its number of enhancers (i.e., Duboule and Wilkins 1998; Levine and Tjian 2003). Enhancers are short DNA sequences; their size is currently evaluated between 100 bp and 1 kb. They were initially described because of their capacity to enhance gene transcription (Banerji et al. 1981). They are enriched for transcription factor binding sites (TFBS), which allow them to respond to the cellular environment (reviewed in Long et al. 2016).

The modular aspects of pleiotropic regulations facilitate phenotypical evolution because it is easier to increase the number of functions of a gene than to create a completely new transcriptional unit (see de Laat and Duboule 2013). Moreover, modifications in regulatory elements affect one pattern of expression without disturbing the others and are often compensated by redundant elements, which buffers the potential negative effects on organism viability. Conversely, mutations in coding sequences affect all functions of the gene, and thus have pleiotropic effects, which tend to be deleterious. The loss of limbs in

snakes and the degeneration of the eye in subterranean mammals illustrate well how regulatory alterations underlie phenotypic changes in natural populations. Regulatory elements that diverged specifically in snakes were associated to genes involved in limb patterning, and those that changed in subterranean mammals were linked to genes relevant for eye development (Roscito et al. 2018).

Evolution of regulatory functions can emerge by repurposing previously existing elements, by exaptation of non-regulatory DNA sequences, often transposable elements, and finally by CRE rewiring upon structural changes (reviewed in Long et al. 2016; Maeso and Tena 2016).

Regulatory landscapes and synteny

If gene multifunctionality and regulation *via* multiple enhancer sequences are not specific to vertebrates, they were never observed to such levels in other species. Particularly in gnathostomes, highly pleiotropic developmental genes are flanked by 'regulatory landscapes' (Spitz et al. 2003), which are large, gene-poor regions where collections of CREs orchestrate pleiotropic transcription patterns (reviewed in Bolt and Duboule 2020). These regulatory landscapes present peculiar genomic features: their origin is more ancient than gene deserts that are not associated with developmental genes, they contain less repetitive sequences and they often correspond to syntenic clusters of conserved non-coding elements (CNEs) (Ovcharenko et al. 2005). The multiplication of regulatory inputs in gnathostomes was accompanied by the expansion of gene regulatory landscapes around developmental genes, sometimes to more than 1 Mb of linear distance. Hence, enhancers are able to modulate transcription from remote genomic locations relative to the promoter of their target genes, a process referred to as "long-range regulation" (Long et al. 2016).

As enhancers generally exert their function by physically contacting their target promoter, they are brought in proximity *via* DNA loops (see de Laat and Duboule 2013). The advent of Chromosome Conformation Capture and single-cell imaging technics have contributed to characterize general features of nuclear organization (reviewed in Cardozo Gizzi et al. 2020; Misteli and Finn 2021; Jerkovic and Cavalli 2021). These methods provided insights that helped formulating an evolutionary scenario to explain the cause and mechanisms of regulatory landscape formation in vertebrates, integrating nuclear organization to the evolution of gene regulation.

Chromatin architecture

At the scale of tens to hundreds of kb, vertebrate genomes are subdivided into topologically associating domains (TADs), which are domains where interactions between sequences occur preferentially (Dixon et al. 2012; Nora et al. 2012; Rao et al. 2014). They emerge from the average frequencies of interactions among a cell population, often of several million cells, in which each cell may have a completely different set of loops, and where the loops themselves are highly dynamic. Hence, TADs are a probabilistic representation of average conformations, and not rigid, permanent structures and consequently, rare fluctuations of chromatin conformations cannot be identified by bulk-assays. Likewise, TAD boundaries are not strict physical borders, but regions where the probability of specific directional interactions drops drastically. As TADs are essentially data structures, sub-configurations and boundaries may appear within big TADs when sequencing resolution increases.

Despite the aforementioned limitations of data collected from bulk-assays, TADs around developmental genes reflect a biological reality that was confirmed by imaging (Cardozo Gizzi et al. 2020). These TADs are generally well delineated by boundaries, which themselves contain gene promoters and binding sites for architectural proteins, primarily CTCF (reviewed in Finn and Misteli 2019). Remarkably, their boundaries often coincide with those of regulatory landscapes, supporting the notion that the range of action of an enhancer appears to be limited to the topological domain it belongs to (Shen et al. 2012; Symmons et al. 2014). Moreover, about 50% of TAD boundaries are conserved between mouse and human (Rao et al. 2014), with some of them being even more deeply conserved (Vietri Rudan et al. 2015). Finally, TADs around developmental genes also correspond to syntenic arrays of non-coding elements (CNEs) (Harmston et al. 2017). For these reasons, TADs have been proposed to form the basic regulatory units of the genome, as they would favour enhancer-promoter contacts located within the same TAD and reduce spurious interactions beyond topological boundaries.

Several genomic loci of high epistemic value were used to investigate the possible functions of TADs. For example, they may provide regulatory robustness by maximizing expression levels, as shown with *Shh* and its ZRS enhancer (Paliou et al. 2019), or enhancer specificity, illustrated by the comparison of forelimb versus hindlimb regulation at the *Pitx1* locus (Kragesteen et al. 2018). In addition, a recent study modifying the conformation of one of the *HoxD* regulatory landscapes related TADs to precise timing of gene activation

(Rodríguez-Carballo et al. 2020). Nevertheless, the role of chromatin folding in gene regulation remains an on-going point of debate (Furlong and Levine 2018; Mir et al. 2019; Oudelaar and Higgs 2021). It is not yet solved whether chromatin structure is encoded independently and is instrumental to gene regulation, or if it reflects dynamic transcriptional processes occurring within nuclei, simply correlating with the various activities of the genome. As an example, disturbances to the 3D conformation of the chromatin produce highly variable effects between loci. In some instances, structural variants reallocate enhancer-promoter contacts, leading to gene misexpression and consequentially produces phenotypes such as malformations (Lupiáñez et al. 2015; Kragesteen et al. 2018; Paliou et al. 2019; Bolt et al. 2021) or cancers (Hnisz et al. 2016; Du et al. 2021, reviewed in (Anania and Lupiáñez 2020; Ibrahim and Mundlos 2020). Yet, in other cases, the effects of TAD disruption on gene expression is mild and has no phenotype (Rodríguez-Carballo et al. 2019, discussed in Ghavi-Helm 2020). In addition, genome-wide impairments of TAD structures in ES cells, through depletion of architectural proteins disturb the activity of only a minority of genes (Nora et al. 2017; Rao et al. 2017; Schwarzer et al. 2017).

The discrepancy in these results raises a number of questions concerning the causality of TAD stability over evolutionary time. A possible explanation could be that TADs have been progressively shaped during evolution, probably because of the adaptive value of placing enhancers close to their target genes, in a genomic environment suited for regulatory activities (i.e., chromatin accessibility, presence of transcription factors and transcription machinery) (reviewed in Bolt and Duboule 2020). In turn, these TADs would provide a genomic environment facilitating the evolution of new regulatory relationships by exposing sequences to promoters and transcription factors, which increases the adaptive value of the system by generating redundancy, compensatory mechanisms and quantitative controls. In this hypothesis, TADs are considered as positive feedback evolutionary systems: the more regulatory elements there are in the loop, the more it will be stabilized and facilitate the emergence of further enhancers. The utilization of a pre-existing architecture for the evolution of new regulation is referred to as “hijacking” of a chromatin architecture (Gonzalez et al. 2007; Schep et al. 2016). This hypothesis is supported by the formation of regulatory landscapes containing a high concentration of CREs and their expansion specifically in vertebrate species, where there are fewer constraints on the total size of the genome compared with rapidly reproducing species with fewer cell types. It is easier to evolve an enhancer in a structure that optimizes the function of another enhancer, than to form the whole structure from the start.

This scenario is also coherent with the 2R-hypothesis. A limitation of the positive feedback loop hypothesis is the accumulation of internal constraints, due to multiplication of regulatory influences (Darbellay et al. 2019; Kentepozidou et al. 2020). At a certain point, this would lead to an evolutionary plateau, until genome duplication releases these constraints by buffering the potentially negative effects of enhancer accumulation. Indeed, the 2WGDs were followed by a great complexification of regulatory modalities in vertebrates (Darbellay and Duboule 2016). Moreover, TADs are tolerant to size changes: they can harbour conserved borders and syntenic arrangements, while their DNA content and amount of regulatory information can greatly vary (Acemel et al. 2017; Bolt and Duboule 2020).

Therefore, it was proposed that TADs provide genomic niches in which the evolution of new regulatory relationships would be favoured (Maeso and Tena 2016; Acemel et al. 2017).

Changes in the nature of developmental gene regulation account for more phenotypical variability than modifications affecting the genes themselves. Therefore, understanding CREs and the structural evolution of regulatory landscapes is essential to explain phenotypical diversification of vertebrates.

Hox GENES TO STUDY REGULATORY MECHANISMS AND INNOVATIONS

Evolution is intimately linked to development. Organisms are shaped during embryonic development and hence, variations in the activities of developmental genes constitute a molecular source of evolution. Genetic changes during embryogenesis produce a diversity of possible phenotypes, which are subsequently kept or lost.

The field of evolutionary biology describes the history of animal transformations and hence, it is not possible to study the evolution by modifying it, as we usually do in genetics. A way around it is to address the source of morphological variations by looking at molecular perturbations during embryonic development, which in a sense recapitulates morphology-producing aspects of evolution. For over 40 years, the study of *Hox* genes has provided important contributions to different fields of research. They first exposed the genetic origin of morphology, then the conservation of master genes across vertebrate and invertebrate species. They also revealed that the orientation and the correct arrangement of genes on the chromosome, in and within clusters, is important for their proper function. In vertebrates, it was shown that *Hox* genes not only build the body plan of different animals, but also establish the regional identities of different structures. In vertebrates it was first shown in the

limbs, then external genitalia and subsequently in many tissues such as the caecum, the kidneys or skin derivatives. These observations helped uncover the pleiotropic aspect of developmental genes, which next led to the description of the modular characteristics of gene regulation, tightly connected to long-range regulation and chromatin architecture. They illustrate well how variations in the regulation of developmental genes contribute to biological diversity more often than modifications in the coding sequences themselves (Edwards et al. 2013). More recently, *Hox* genes were useful to study how chromatin topology may influence or organize gene transcription.

Hox genes encode transcription factors, modulating the activities of downstream genes by binding specific DNA sequences. They belong to the ANTP class of the homeobox gene family, which is characterized by a deeply conserved sequence of 180 bp called the "homeobox" (Fjose et al. 1985). The homeobox produces a 60 amino acid homeodomain, constituted of three alpha helixes. Helixes 2 and 3 form a protein motif called helix-turn-helix, which can bind DNA. Helix 3 binds in the major groove of the DNA and achieves sequence recognition (Galliot et al. 1999). The initial function of *HOX* proteins in patterning the main body axis is conserved among Bilaterians (reviewed in Garcia-Fernàndez 2005). They form a molecular code, providing positional identities to the cells of a growing embryo, working like a system of coordinates. In vertebrates, the position of *Hox* transcripts along the embryonic trunk delimits the regional identities of the future axial skeleton. For example, the members of the *Hox* family specify the levels at which ribs should grow, and overexpression of these genes induce ectopic ribs at cervical and lumbar levels (Krumlauf 1994; Duboule 1996; Mallo et al. 2010). More generally, the anterior boundaries of *Hox* gene expression can be transposed to the morphological boundaries of vertebrae.

Hox **genes in flies and in vertebrates – conservation beyond sequence**

Hox genes were discovered in the fruit fly. In 1978, Ed Lewis observed in his stocks of mutant flies that some of them had their antennas replaced by a pair of legs. He predicted a genetic cause for these malformations, proposing the existence of a gene family that would be in charge of organizing the segmental identities along the antero-posterior (A-P) axis (Lewis 1978). This hypothesis was confirmed few years later by the isolation and characterization of the DNA fragment responsible for these malformations (Bender et al. 1983; Garber et al. 1983). Most remarkably, homologous genes were found in the mouse (McGinnis et al. 1984) and in human (Levine et al. 1984). In addition to their conservation at the sequence level,

Hox gene organization in clusters was shown to also be conserved in mouse (Duboule et al. 1986; Duboule and Dollé 1989), as well as their function as developmental regulators involved in segment patterning (Gaunt et al. 1986). A striking genomic feature is observed and termed 'spatial collinearity' (Lewis 1978; Gaunt, Duboule 1988; Kondo et al. 1998), in which the order of the genes on the chromosome reflects the distribution of their transcription patterns along the A-P axis. In animals that develop sequentially from head to tail, a 'temporal collinearity' is observed, meaning that the sequential activation of *Hox* genes also present a temporal dimension (extensive review about collinearities in Noordermeer and Duboule 2013).

These results, showing that mouse and fly have *Hox* clusters with colinear transcriptional activation and major roles in body plan patterning, revealed the structural and functional conservation of *Hox* clusters between invertebrates and vertebrates. At the time, it was remarkable to conclude, that, at the end, genes and pathways organizing body plans are fundamentally the same in insects and in mammals (Bachiller et al. 1994).

Hox **gene evolution in Metazoa**

Hox genes probably originated early in animal evolution, in a common ancestor of cnidarians and bilaterians (600 MYA), although their presence in poriferans is still under debate (Pastrana et al. 2019). Gene duplication in *cis* likely generated a *protoHox* cluster and dating the first *Hox* genes consist in determining when the duplication of this ancestral *protoHox* cluster produced a *Hox* cluster distinct from a *paraHox* cluster. In 1992, they were identified for the first time in a cnidarian species, *Chlorohydra viridissima,* where they were shown to be involved in regenerative processes (Schummer et al. 1992). Whether the first *Hox* genes already functioned as patterning factors in the ancestor of cnidarians and bilaterians remains an open question (Reddy et al. 2015; Rentzsch and Holstein 2018). In the sea anemone *Nematostella vectensis,* they are expressed in ordered domains along the directive axis of the larva (Martindale 2004), suggesting a link between segmentation and *Hox* expression. In another study, the function of *Hox* genes in the sea anemone was addressed by genetically abrogating their sequences or their expression, which resulted in altered endoderm segmentation and some defects in tentacle patterning (He et al. 2018). If these two studies do demonstrate an association between *Hox* genes and segment patterning in *Nematostella vectensis*, one should be cautious inferring this function to ancestral *Hox* genes. Indeed, extant cnidarian species are highly derived versions of their ancestors, and

research is limited to a few model organisms, namely hydra and *Nematostella vectensis*. Hence, convergent evolution cannot yet be ruled out with parsimony.

Pleiotropy leads to increased cluster organization

Following the genesis of the ancestral cluster, additional tandem duplications increased again the number of *Hox* genes per cluster (Garcia-Fernàndez 2005; Holland 2013). In deuterostomes, a maximum of fifteen *Hox* genes have been reported so far, in *Amphioxus* (Lemons and McGinnis 2006; Holland et al. 2008; Amemiya et al. 2008). The *Hox* cluster of *Amphioxus* is often used to illustrate the ancestral situation (Marlétaz et al. 2018). According to the 2R-hypothesis, jawed vertebrates (gnathostomes) present four clusters, *HoxA* to *HoxD,* descending from a single locus of the chordate ancestor (**Figure 2**). The genes of each cluster are classified in 13 paralog groups, determined by their relative position and sequence similarities.

The evolutionary origin of *Hox* clusters in cyclostomes (Lampreys, Hagfishes, Chimeras) is challenging and has long been uncertain. Recently, the reconstruction of a proto-vertebrate and a proto-cyclostome genomes supported the scenario of a first WGD (1R) in the vertebrate ancestor, followed by a cyclostome-specific hexaploidization (triplication of the genome) on one side, and a gnathostome-specific second duplication on the other side (2R) (Nakatani et al. 2021). In addition, a third round of WGD (3R) happened at the base of the teleost lineage (**Figure 2**).

Duplications of *Hox* clusters probably opened a window of opportunities for the establishment of new regulatory relationships, controlling their expression in vertebrate-specific structures (Soshnikova et al. 2013). Pioneer studies showed that in addition to their role in patterning the main body axis, *Hox* genes were activated again in the limb to provide positional information (Dollé et al. 1989; Oliver et al. 1989; reviewed in Duboule 1991a). Moreover, their expression in the limb is conserved in chicken, where the association between *Hox* transcription patterns and morphology was shown early on. Manipulations of chicken wing buds either by grafting cells from the polarizing region or applying retinoic acid showed that when anterior cells form posterior structures, the expression of *Hox* genes is changed accordingly and becomes posterior-like (Duboule 1991b). Some genetic evidences supported this previous observation, showing that absence of polarity in wings correlates with altered *Hox* transcription domains (Izpisúa-Belmonte et al. 1992). These studies showed that

modifications of *Hox* gene expression can be associated to morphological changes in other structures than the main body axis (Zákány et al. 1997; Gérard et al. 1997).

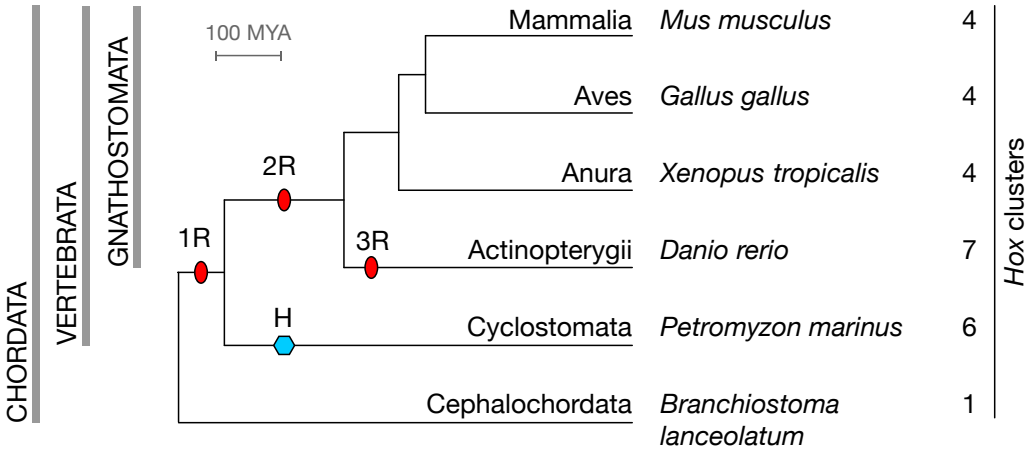


Figure 2 - Evolution of *Hox* clusters in vertebrates.

Phylogenetic tree showing major classes of Chordata, class names are indicated on the branches together with the name of a representative species and the number of *Hox* clusters found in each respective genome. The current classification of Chordata includes Cephalochordata (Amphioxus), Tunicata (Sea squirts) and Vertebrata. Red ovals represent whole genome duplication events and the blue hexagon a whole genome triplication event (hexapolyploidization). According to our current knowledge of vertebrate phylogeny and genomic evolution, a first round (1R) of WGD occurred at the base of the vertebrate lineage, preceding the divergence of Cyclostomata from what will form the Gnathostomata lineage. Following this divergence, and a whole genome triplication event occurred at the base of Cyclostomata, independently from a second round of WGD (2R) early during the evolution of Gnathostomata. These events would have produced, from the unique *Hox* cluster of the chordate ancestor, six complexes in the genomes of Cyclostomata, and four in those of most jawed vertebrates. In addition, Actinopterygii would have undergone an additional round of WGD (3R), followed by a loss of one *Hox* cluster, explaining the 7 complexes found in the genome of *Danio rerio*.

Hox gene pleiotropy may have contributed to the rapid diversification and complexification of jawed vertebrates. Paralleling the multiplication of their functions, the *Hox* clusters underwent peculiar evolution towards organization, discussed in (Darbellay and Duboule 2016). They are compact, with few repeated sequences, no unrelated genes, and all genes are transcribed from the same strand of DNA. This characteristic cluster consolidation is a genomic synapomorphy of jawed vertebrates (Duboule 2007), that may have optimized the transcriptional response to accumulating enhancers in the regulatory landscapes (for review and references, see Darbellay and Duboule 2016; Hrycaj and Wellik 2016).

The mouse *HoxD* locus

The collinear regulation of *Hox* genes along the main body axis represents their initial function and is kept at the four clusters, where it is established by local regulatory elements. By contrast, their expression patterns in vertebrate specific structures rely on regulatory influences located outside of the clusters, probably to prevent interferences with the more ancestral regulation. The mouse *HoxD* locus is a paradigm of how highly pleiotropic gene regulation is organized in the genome. The *HoxD* locus is a cluster of nine genes covering approximately 100kb of linear genomic distance. It is flanked on each side by two large (~1Mb) regulatory landscapes of distinct and complementary activities. While bacterial artificial chromosome (BAC) transgenes carrying the cluster alone recapitulate the collinear patterns of *Hoxd* genes along the trunk, they do not reproduce those in more recently evolved structures (Spitz et al. 2001; Guerreiro et al. 2016) such as limbs (Dollé et al. 1989; Hérault et al. 1998), external genitalia (Dollé et al. 1991) or cecum (Zákány and Duboule 1999). By contrast, *Hoxd* genes in novel vertebrate structures or organs are under the control of long-range regulatory elements located in the regulatory landscapes. The telomeric domain (T-DOM) is located 3' to the cluster and regulates *Hoxd* genes in many different structures, as shown by series of genetic modifications (Spitz et al. 2005). On the 5' side of the cluster, the centromeric domain (C-DOM) was shown to activate *Hoxd* genes in digits and in genital tubercles. (Montavon et al. 2011; Andrey et al. 2013; Lonfat et al. 2014). It was observed that the C-DOM enhancers activating *Hoxd* genes in digits are constitutively found close to their promoter in the 3D-space, meaning even in tissues where *Hoxd* genes are not expressed (Montavon et al. 2011). These results suggested the existence of a poised conformation of regulatory landscapes, which was subsequently confirmed by the characterization of TADs.

The *HoxD* cluster actually coincides with the TAD boundary separating the C-DOM from the T-DOM. These two TADs seem to partition regulatory activities as they were so far never seen to be active at the same time, in the same tissue (Andrey et al. 2013; Rodríguez-Carballo et al. 2017). Moreover, if the active contribution of TADs is still under debate, recent results using the mouse *HoxD* cluster have been integrated in the conceptual framework describing how their constitutive conformation could serve as niche to evolve new enhancers.

In summary, duplications of *Hox* clusters created opportunities for the establishment of new regulatory relationships, controlling their expression in vertebrate-specific structures (Soshnikova et al. 2013). It resulted in increased numbers of regulatory elements, found further from their promoter and hence a more important role for chromatin architecture. Recent results using the mouse *HoxD* cluster have been organized in a conceptual framework integrating the evolution of vertebrate genomic mechanisms in the context of development. The complexification of regulatory mechanisms at developmental genes parallels the morphological complexification of animal species.

SCOPE OF THE THESIS

According to our current knowledge, the emergence of large regulatory landscapes around *Hox* complexes, the consolidation of *Hox* clusters and the multiplication of *Hox* gene functions all followed the 2 WGDs that occurred at the base of the vertebrate lineage and paralleled the evolution of anatomical synapomorphies in gnathostomes. These observations were integrated into a coherent scenario by (Darbellay and Duboule 2016) to try to explain the evolution of *Hox* clusters. In this proposal, the multiplication of *Hox* clusters released the internal constraints of the ancestral *Hox* cluster (*Hox-Anc*) by generating redundant and compensatory genetic mechanisms, which opened a window of opportunities for the evolution of new gene regulations and associated morphological innovations. Constitutive chromatin domains form genomic environments where the evolution of productive regulatory interactions is favoured. In turn, the accumulation of enhancers in regulatory landscapes tightens the TAD structure and simultaneously, the adaptive value of optimizing interactions between *Hox* promoters and remote enhancers favour cluster compaction.

The amount of genomic information collected since then has not proved this scenario wrong. However, it still remains a conceptual framework, lacking horizontal comparisons of the structures and functions of different *Hox* loci. Understanding the mechanistic links that

could explain why the association of TADs, regulatory landscapes and pleiotropy at vertebrate *Hox* loci was conserved over millions of years of evolution will be important for our understanding of regulatory evolution and associated phenotypic diversity of vertebrates.

In this work, we used the *HoxD* cluster as system to address the relationships between chromatin architecture and regulatory activities in the evolution of *Hoxd* gene regulation. We used a combination of *in silico* comparative analyses and experimental approaches, including genetic assays, biochemical analyses and gene transcription tests to establish detailed comparisons of *HoxD* loci, focusing on mouse and chicken.

Our aim is to better understand the contribution of chromatin architecture in the emergence of new regulatory relationships between vertebrate developmental genes and their associated long-rang enhancers.

SECTION I

EVOLUTION OF GENOMIC ORGANIZATION AT VERTEBRATE *HOX* CLUSTERS

INTRODUCTION

The *Hox* clusters of jawed vertebrates systematically display a tighter organization than their invertebrate counterparts. Various comparisons of metazoan genomes have demonstrated that the most parsimonious scenario to explain the unique features of vertebrate *Hox* clusters is an evolutionary force that selects towards organization (Duboule 2007). In general, gene clusters are associated with global regulatory strategies, which are thought to simplify the coordination of gene activation. Compaction would have the effect of likely optimizing the transcriptional response to a variety of long-range enhancers distributed within large regulatory landscapes (Darbellay et al. 2019). In regard to the fact that *Hox* genes were co-opted during the evolution of multiple vertebrate synapomorphies (i.e., Spitz et al. 2005), it was proposed that a functional relationship exists between the genomic organization of *Hox* clusters and the evolution of novel regulatory inputs. In addition, the regulatory landscapes controlling the expression of pleiotropic developmental genes generally correspond to large and well-delineated TADs (Harmston et al. 2017). This raises the interesting possibility that TADs embedding regulatory landscapes can display distinct properties where potentially the 3D organization of regulatory landscapes is interconnected with the evolution of pleiotropy.

The mouse *HoxD* cluster is particularly well organized (i.e., it is condensed, excludes most non-*Hox* genes and all *Hox* transcription units are completely polarized) (**Figure 3A**) and sits at the interface between two regulatory landscapes, the C-DOM in 5' and the T-DOM in 3'. Moreover, each landscape is encompassed within its own Topologically Associating Domains (TADs), the 5'-TAD and the 3'-TAD (**Figure 3B**).

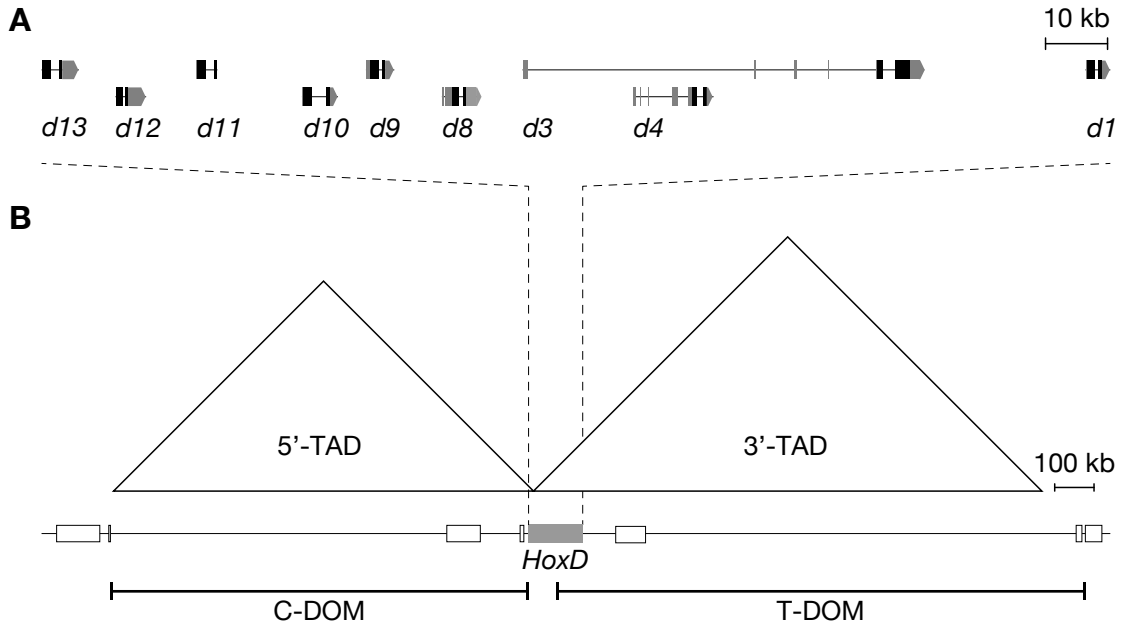


Figure 3 - The genomic structure of the mouse *HoxD* locus.

A A detailed map of the *HoxD* gene cluster (chr2:74667015-74768000) is shown. Gene names were abbreviated from '*Hoxd'* to '*d*' and are annotated at the below the track. **B** schematic representation of the *HoxD* locus (chr2:73800000-75800000) showing the presence of TADs (triangles) corresponding to the span of regulatory landscapes (scale bars) at the *HoxD* locus. Genes are symbolized by white rectangles, and the *Hoxd* genes as a single grey rectangle. TAD calling from (Beccari et al. 2020). **A,B** Dashed lines indicate the span of the *HoxD* cluster.

In addition, murine *Hoxd* genes are expressed in different anatomical structures of different evolutionary origins (**Figure 4A,B**). For example, somites are found in Amphioxus, neural crest cells (NCC) derived structures in lampreys, and legs in tetrapods (Shu et al. 2003; York et al. 2020; Wood and Nakamura 2018). *Hoxd* gene expression along the major body axis represents their ancestral function and relies on regulatory elements mostly located within the *HoxD* cluster, as shown by the X-gal staining of embryos transgenic for a BAC that covers the *HoxD* cluster (**Figure 4C**, BAC^{HoxD}). In contrast, *Hox* gene transcription in more recently evolved structures depend on enhancers localized in the regulatory landscapes. For example, X-gal activity was detected in vibrissae primordia of embryos carrying BAC^{Mtx2} and in limbs of those transgenic for BAC^{T2}. In turn, different patterns in facial cells deriving from NCC where detected in each of the BACs spanning the T-DOM (**Figure 4C**, BAC^{Mtx2}, BAC^{T1,} BAC^{T2}). For these reasons, the mouse *HoxD* locus represents a particularly interesting genomic system to address the relationships between chromatin organization and the evolution of gene regulation.

One way to evaluate the relationship between the transcriptional control of *Hoxd* genes and the tight genomic topology of this locus (cluster and regulatory landscapes) is to evaluate the relative compactness of different *Hox* gene clusters, as well as the relative expansion of the adjacent regulatory domains. We did so by comparing the mouse *HoxD* locus, first to the *HoxD* locus of several chordate species, and then comparing to the *HoxA*, *HoxB*, and *HoxC* loci in mouse. In these analyses, the C-DOM is defined as the region spanning from *Atp5g3* to *Hoxd13,* the *HoxD* cluster as the distance between *Hoxd13* and *Hoxd4*, and the T-DOM as the region from *Hoxd4* to *Nfe2l2.* To make comparative measurements of the distances between corresponding regions, in different species or at paralogous mouse genomic loci, the transcription start site (TSS) of homologous genes were used as anchor points.

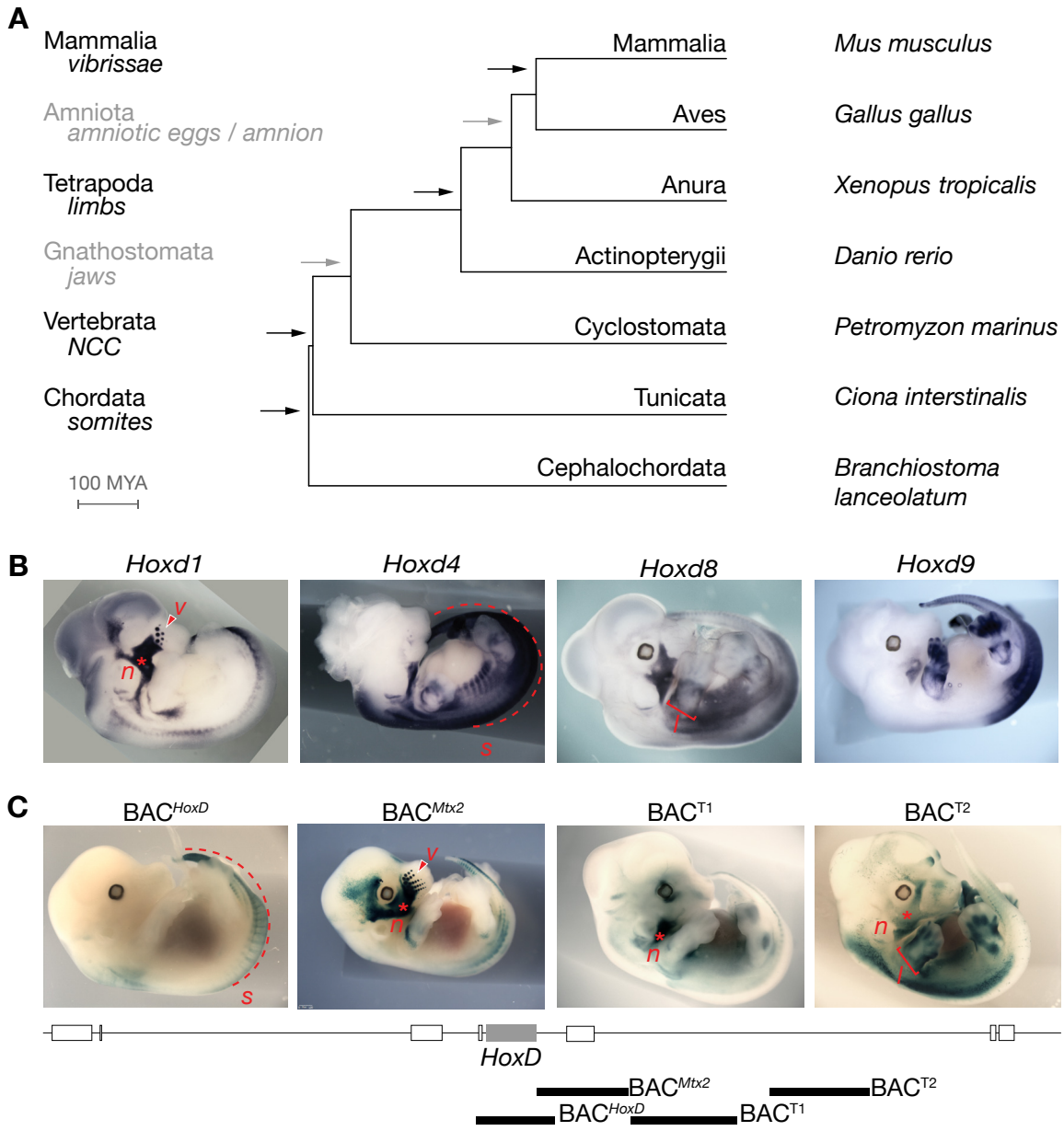


Figure 4 - Emergence of vertebrate synapomorphies.

Figure 4 - Emergence of vertebrate synapomorphies.

A Evolutionary tree of Chordata. Phylogenetic groups are indicated in grey on the left of the tree, together with synapomorphies (*italics*) that likely were determinant in the evolution and diversification of vertebrate species. Synapomorphies for which *Hoxd* gene expression is shown in the panel 'B' are in black font, other representative examples are in grey font (jaws, amnion). Arrows indicate the inferred emergence of each group and associated feature. Class names are indicated on the branches, and on the right of the tree, a representative species that was used in our analyses is shown. All chordates have somites and a notochord at some point of their development. Cyclostomata are represented by hagfishes and lampreys and are important milestones in vertebrate evolution as they mark the emergence of tissues derived from the neural crest cells (*NCC*) (numerous cell types that uniquely define vertebrates; among others dentine, the head cartilage, skeletal cells and connective tissue). They diverged from Gnathostomata (jawed vertebrates), which groups Actinopterygii (ray-finned fishes, i.e., sturgeon, gar), Chondrichthyes (sharks, chimeras and rays, not shown) and Tetrapoda. Tetrapoda are characterized by a pair of legs and were probably pioneer animal species that colonized terrestrial environments. The emergence of a protective membrane around the embryo, the amnion, accompanied a reproductive mode released from the need of water and characterize the Amniota. Finally, the evolution of lactation is a synapomorphy of Mammalia. **B** Whole mount *in situ* hybridization (WISH) on E12.5 mouse embryos showing different expression domains of several *Hoxd* genes. Probe names are indicated on the top of the pictures. **B,C** The expression of *Hoxd* genes are indicated in red as follows: asterisk for neural crest cell derivatives (*n*), arrowhead for vibrissae primordia (*v*), dashed line for somites (*s*) and brackets for limbs (*l*). The patterns of expression are indicated only on the embryo where they are the most visible, but they may belong to several *Hoxd* genes. **C** X-gal staining on E12.5 mouse embryos carrying randomly integrated BACs. A scheme of the mouse *HoxD* locus (chr2:73800000-75800000) is represented below, genes are shown as empty rectangles, the genes of the *HoxD* cluster as a single grey rectangle and BAC transgenes are indicated as thick lines.

RESULTS

Evolution of vertebrate *HoxD* clusters

We first compared the mouse *HoxD* locus to chicken (*Gallus gallus*), frog (*Xenopus tropicalis)* and zebrafish (*Danio rerio*) orthologs. We included both the lamprey *Hoxδ* (basal vertebrate) and the Amphioxus *Hox* loci (chordate) to consider a complete evolutionary perspective, from the root of the vertebrate lineage through the more recently derived tetrapods. This allows for an analysis of the *HoxD* locus in several species containing shared and unique morphological characteristics of major importance in the evolution of vertebrates.

To compare the level of cluster compactness relative to regulatory domain expansion in different species, we first determined the T-DOM homolog and the *HoxD* cluster homolog in each species. We excluded the C-DOM from this analysis as both the amphioxus and the lamprey genomes do not show unambiguous homology with murine *Atpg5g3* (**Figure 5A**). To measure the level of T-DOM expansion relative to cluster compactness, we divided the genomic size of the T-DOM by the size of the *HoxD* cluster for each species. To detect a possible evolutionary trend in the resulting ratios, species were ordered on the *x-axis* by time of independent evolution relative to mouse, in million years ago (MYA) (**Figure 5B**). For the two species with the most ancient divergence time with mice, Amphioxus and Lamprey, the length of the T-DOM is approximately the same size as this of the *Hox* cluster and the *Hoxδ* cluster, respectively. By contrast, the size of the T-DOM expands progressively compared to the size of the cluster in gnathostomes, correlating with the emergence of synapomorphies in which *Hoxd* genes were likely co-opted.

Cluster compactness or regulatory expansion or a combination of both can account for the observed ratio increase. To determine what is the main driver of this augmentation, the size of each domain and cluster was normalized to genome size and compared to mouse. Despite some variability, the span of regulatory landscapes remains in the range of two-times larger or smaller than in the mouse, indicating that regulatory landscapes generally expand or shrink proportionally to the whole genome (Harmston et al. 2017). In contrast, we found that the *HoxD* gene cluster is systematically more compact in mouse than in other species (**Figure 5C**).

Altogether, these results support the scenario in which the *HoxD* cluster progressively consolidated during gnathostome evolution (Duboule 2007), paralleling the increasing complexity of their regulatory modalities (Spitz et al. 2005).

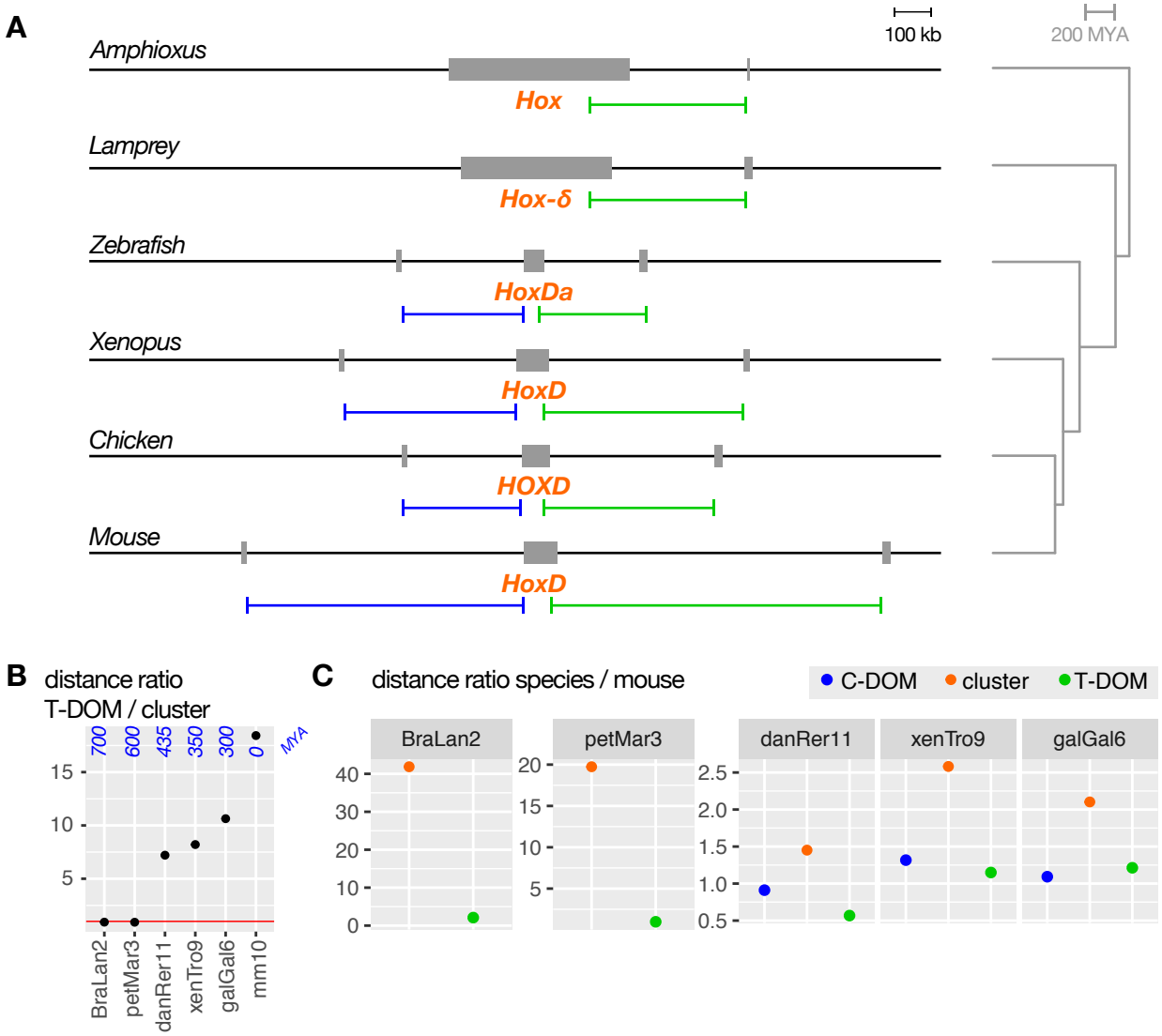


Figure 5 - *HoxD* loci in chordates.

A *Left* Schemes of chordate *Hox* clusters orthologous to mouse *HoxD*. *Right* phylogenetic relationships. **B** *Y-axis* distance ratio T-DOM / cluster. *X-axis* is the genome assembly symbol. *Horizontal red line* indicates the 1 value. *Blue numbers* at the top of the plot show the evolutionary time separating each species from the mouse, in million years ago [MYA]. **C** *Y-axis* distance ratio of a species region / mouse orthologous region. *X-axis* genomic regions. **B,C** The C-DOM is defined as the distance from *Atp5g3* to *Hoxd13,* the cluster as the distance from *Hoxd13* to *Hoxd4* and the T-DOM as the distance from *Hoxd4* to *Nfe2l2*. The reference is the mouse and homologous genes are used to define corresponding regions in the other species.

Organization of the four mouse *Hox* clusters

Our previous analysis suggested a functional link between the tight organization of *Hox* clusters in jawed vertebrates and their high level of pleiotropy during embryonic development. By contrast, the sizes of regulatory landscapes appeared to be more related to the total genome size than to the multiplicity of *Hoxd* gene expression patterns. Previously, it has been shown that TAD size scales with genome size (Harmston et al. 2017). Nevertheless, differences in genome sizes can result from many other reasons than the level of gene pleiotropy (transposable elements, repetitive sequences, copy number variation, polyploidization) and thus, normalization to total genome size may induce biases. To circumvent this limitation, we compared the sizes of clusters and regulatory landscapes at the four mouse *Hox* loci. The functions of these four loci do not completely overlap, each having some discrete pleiotropies, so while they have been maintained within the same genomes, and they have each acquired approximately 600 million years of divergence, they are not under the same selective forces. If there is an advantage to cluster compaction or landscape expansion, it may be visible at some *Hox* loci and not at others.

We used the three segments previously defined at *HoxD* as C-DOM, cluster and T-DOM to determine the paralog segments at *HoxA, HoxB* and *HoxC.* Published Hi-C data was used to visually compare the global interaction profiles of each locus and to determine the approximate positions of the chromatin compartments where the clusters are found.

Strikingly, the basic organization of the *HoxA* landscape is very similar to *HoxD* in regard to the fact that *HoxA* is positioned at the boundary between two TADs encompassing regulatory landscapes, and well isolated from each other (Berlivet et al. 2013; Woltering et al. 2014; Gentile et al. 2019) (**Figure 6A**). This peculiar chromatin architecture is not observed at *HoxB* and *HoxC.* Even though the *HoxC* cluster is also flanked by two domains of frequent interactions, and which correspond to the paralog T-DOM and C-DOM, these domains are considerably smaller than those of *HoxA* and *HoxD*. Moreover, they are not well insulated from each other as indicated by the single, larger domain in which they are included, but for which no TAD was called. At the *HoxB* locus, a single TAD is called with limits corresponding to those of the *HoxB* regulatory landscapes.

We next measured the level of cluster compactness relative to the expansion of regulatory landscapes by dividing the genomic size of each paralog T-DOM by the genomic size of each paralog *Hox* cluster. As there is no *Atp5g* paralog at the *HoxA* locus, TAD sizes were compared between *HoxA* and *HoxD* as a proxy for gene regulatory deserts size but not for *HoxB* and *HoxC* as these clusters do not share this common chromatin architecture

(**Figure 6A**) (Woltering et al. 2014). The ratios we obtained were clearly higher at *HoxA* and *HoxD*, where the size of the T-DOM is more than 10 times larger than the cluster. By contrast, the *HoxB* and *HoxC* size ratios were below 5 (**Figure 6B**). We finally evaluated whether the differences of ratios between the different *Hox* loci are due to more compact cluster, wider regulatory landscapes, or a combination of both. The size of each segment at *HoxA, HoxB,* and *HoxC* was reported to the size of the corresponding *HoxD* segment. In the mouse, the *HoxA* and *HoxD* clusters are more compact than *HoxB* and *HoxC*, and their regulatory landscapes are larger (**Figure 6C**).

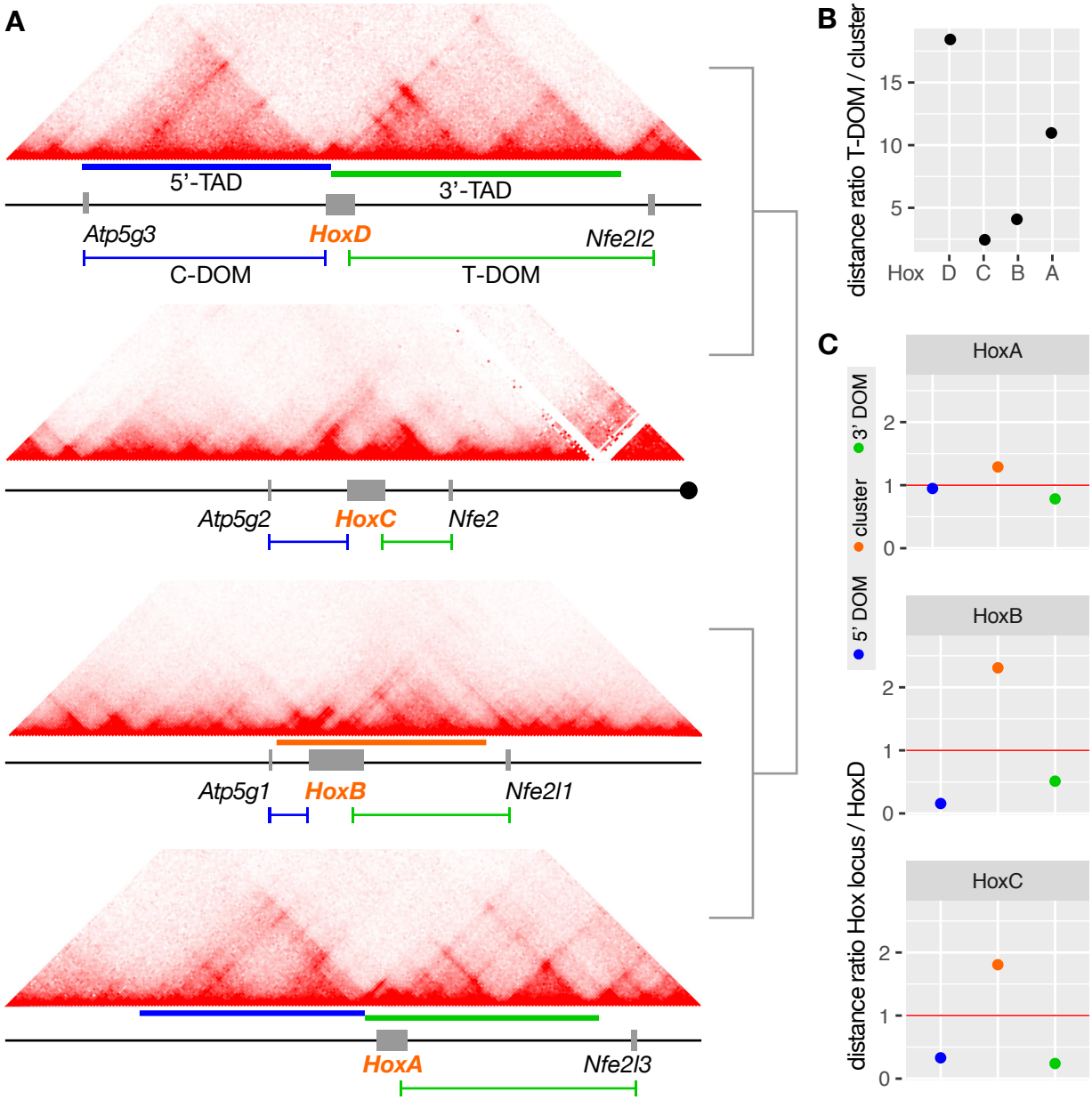


Figure 6 - The four paralogous mouse *Hox* loci.

A *HiC heatmap* 10kb bin resolution, mESC, data from (Bonev et al. 2017), TAD calling from (Beccari et al. 2020) *blue* 5'-TAD *green* 3'-TAD, *grey boxes* are genes, *blue and green scale* represent the C-DOM and T-DOM respectively, as used for quantifications in B,C. The *black dot* represents the end of chromosome 15 (*HoxC*). The dendogram on the right shows the evolutionary relationships between *Hox* clusters, as established by protein homology in (Ravi et al. 2009) **B** *Y-axis* distance ratio T-DOM / cluster. **C** Distance ratio of a region divided by the paralogous region at the *HoxD* locus. **B, C** The C-DOM is defined in the mouse as the distance from *Atp5g3* to *Hoxd13,* the cluster as the distance from *Hoxd13* to *Hoxd4* and the T-DOM as the distance from *Hoxd4* to *Nfe2l2*. Paralog genes are used to define corresponding regions in the other species. The *HoxA* C-DOM is defined as the size of the 5'-TAD as there is no paralog for *Atpg5g3.*

Our comparative analysis of the four mouse *Hox* complexes revealed that a tighter organization of the cluster (*HoxA* and *HoxD*) systematically correlates with the presence of two wide regulatory landscapes corresponding to two well-insulated TADs. According to our current knowledge, the first round of duplication produced a *HoxA/B* cluster and a *HoxC/D* cluster. Then, a second round of duplication generated the *HoxA, HoxB, HoxC* and *HoxD* clusters (Ravi et al. 2009). Hence, the similar genomic properties of *HoxA* and *HoxD*, or of *HoxB* and *HoxC*, seem to result from convergent evolutionary events.

In light of these observations, we propose that all four *Hox* clusters in mouse evolved in parallel towards cluster consolidation and regulatory landscapes expansion, but to different degrees. Both *HoxA* and *HoxD* loci converged on a similar bipartite organization, while *HoxB* and *HoxC* retained more features of the ancestral *Hox* cluster (*Hox-Anc*), resembling the single 3D compartment and premises of T-DOM of the amphioxus *Hox* cluster (Acemel et al. 2016). Two other arguments support this interpretation as well. First, the regulatory convergence of the *HoxA* and *HoxD* complexes has been established in digits and genital tubercles (Lonfat et al. 2014). Then, the *Hox-Anc* locus has been duplicated before the expansion of regulatory landscapes and before cluster compaction, and hence was very likely organized into a single loose TAD, as inferred by the topology of the amphioxus *Hox* cluster (Acemel et al. 2016).

DISCUSSION

The six non-consolidated lamprey *Hox* clusters

A possible scenario to explain the evolution of *Hox* clusters could be that initially, the multiplication of *Hox* clusters would form a permissive environment for the evolution of new gene regulations and new morphological innovations. However, it would not directly trigger such evolutionary events, as jawless vertebrates and most teleost species have at least six clusters and do not display particularly high levels of cluster compaction and morphological complexity. In gnathostomes, pioneer regulatory contacts between promoters and remote enhancers would have consolidated the ancestral 3D structures of the regulatory landscapes, which in turn, would facilitate the emergence of additional enhancers. *Hox* clusters would become more compact and organized, optimizing their responses to remote enhancers.

Previous studies reported that the Lamprey genome underwent 3 WGDs and thus has at least six *Hox* clusters. However, these clusters have not been consolidated, at least not to a comparable way as in jawed vertebrates (cf., **Figure 2**) (Duboule 2007). In this regard, it is

important to specify that the multiplication of genomic loci may have permitted, and not provoked, the complexification of gene regulation and the evolution of a diversity of vertebrate morphologies, physiologies and behaviours. For example, most jawless vertebrates and teleost fishes have six clusters and do not display higher levels of cluster compactness (Lamprey) or of *Hox* gene pleiotropy (fishes) than other vertebrates. In addition, polyploidization in insects and plants drives evolution (Li et al. 2018) but is not automatically related to complexification. (Duboule 2007). Hence, the release of internal genomic constraint appears to not systematically lead to the implementation of the 'positive evolutionary loop' proposed above, in which remote enhancers consolidate TADs and that TADs help the emergence of new remote enhancers. The establishment of pioneer productive interactions may have not occurred in Lamprey or did not present adaptive values at this stage and were not kept. Alternatively, the processes may have started, but emergence and consolidation of global regulations may occur at much slower rates than in gnathostome genomes. It would be interesting, in the future, to look at chromatin interactions and regulatory activities in different Lamprey loci, for *Hox* clusters and other key regulatory of development (i.e., *Shh*, *Pax*, *Meis*), to see if we can detect some regulatory landscapes, long-range enhancers contacting promoters.

The different evolutionary trajectories of mouse *Hox* clusters

According to the 2R hypothesis and protein sequence comparisons, a first whole genome duplication would have generated *HoxA/B* and a *HoxC/D* cluster from a single ancestral *Hox* cluster (*Hox-Anc*). The second whole genome duplication would have generated a *HoxA* and a *HoxB* cluster from *HoxA/B*, and a *HoxC* and *HoxD* cluster from *HoxC/D* (Ravi et al. 2009), early during vertebrate evolution.

The bimodal chromatin structure of a gene cluster establishing differential contacts with two large regulatory landscapes located on each side characterizes the *HoxA* and *HoxD* clusters but is not observable at the *HoxB* and *HoxC* clusters. It was proposed that the function of this typical topology is to separate the regulation of posterior genes from the others, a necessity resulting from the posterior prevalence (Woltering et al. 2014). In this scenario, the C-DOM and T-DOM were already pre-formed in the *Hox-Anc* cluster, and their absence in *HoxB* and *HoxC* would be due to a loss of this organization, explained by the cluster structure that would be sufficient to isolate posterior genes from the rest of the cluster.

In contrast, we propose that the topological organization of *Hox-Anc* resembled this of Amphioxus, with a less dense *Hox* cluster embedded in a single TAD and some premises of

regulatory landscape in 3' (Acemel et al. 2016) (e.g., small genomic domains with some CREs but also unrelated genes). Following the 2 WGDs, gene multi-functionalization was widespread across vertebrate genomes and thus, the four *Hox* clusters acquired new functions, became better organized and some small regulatory deserts started to form.

We want to propose two non-exclusive scenarios to explain the clear structural similarities between *HoxA* and *HoxD,* and why they are absent from *HoxB* and *HoxC.* A first reason could be, as formulated in (Woltering et al. 2014)*,* the likely existence of an evolutionary force towards a functional separation of posterior *Hox* genes from anterior *Hox* genes, due to the prevalent function of the former on the latter. This adaptive pressure would be stronger at *HoxA* and *HoxD,* where all the genes are closer to each other than at *HoxB* and *HoxC,* where the mere structure of the cluster may be sufficient isolate posterior genes. Alternatively, *Hox* cluster duplications would release the functional constraints on one of the two loci, allowing *Hoxa* and *Hoxd* genes to acquire numerous additional functions. As TAD structure and accumulation of remote regulatory elements appear to lead to some kind of virtuous evolutionary cycle, once the first productive regulatory contacts were established between promoter and pioneer enhancer sequences, the evolutionary trajectories of *HoxA* and *HoxD* would continue to diverge more, or faster, from the *Anc-Hox* than *HoxB* and *HoxC.*

Functional comparisons of gene expression patterns, regulations and interactions

In the interpretation of our results, we use the size of regulatory landscapes as a proxy for the complexity of gene expression (i.e., the number of spatially and temporally discrete transcription domains of a given gene). We based this assumption on the work of (Clément et al. 2020)*,* which shows that the number of CREs regulating a gene correlate with the number of tissues in which this gene is expressed. We thus analyzed the sizes of clusters and regulatory landscapes in different genomes and at different loci, which allows for a multiway comparison across developmental and evolutionary timescales. However, the addition of horizontal comparisons of *Hox* gene transcription patterns would reinforce our conclusion that cluster compactness, size of regulatory landscapes and multiplication of gene functions are correlated. Ideally, the regulatory potential and the interaction profiles of each locus of the C-DOM and T-DOM should be evaluated in each case.

MATERIAL AND METHOD

Phylogeny

Phylogenetic trees were generated from http://www.timetree.org/ and edited with seaview and Illustrator.

Species input for tree in Figure 1: *Chondrilla aff. nucula chond, Nematostella vectensis, Trichoplax, Sepioteuthis australis, Mus musculus*

Species input for tree in Figure 2: *Mus musculus, Gallus gallus, Xenopus laevis, Danio rerio, Petromyzon marinus, Branchiostoma lancealatum*

Comparisons of *Hox* clusters in different species

To compare *Hox* loci in different species, we used the amphioxus (*Branchiostoma lancealatum*) single *Hox* locus, the lamprey (*Petromyzon marinus*) *Hox-∂* locus, the zebrafish (*Danio rerio*) *HoxDa* locus and finally, the *HoxD* clusters of frogs (*Xenopus tropicalis*), chicken (*Gallus gallus*) and mouse (*Mus musculus*).

Amphioxus assembly BraLan2 was visualized in a UCSC Genome Browser track hub from (Marlétaz et al., 2018), ressources avalaible at http://amphiencode.github.io/Data/. The Lamprey genome and *Hox* clusters were evaluated referring to (Smith et al. 2013). Vertebrates *HoxD* loci were found using the convert tool from the UCSC genome browser http://genome.ucsc.edu.

Genome sizes were obtained from https://www.ncbi.nlm.nih.gov/assembly, using "RefSeq Assembly Accession" and searching for "Total sequence length".

Data were plotted using R: "*R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* " https://www.R-project.org/ .

Table 1 – List of *Hox* or *HoxD* loci in different species.

chr	start	stop	Assembly	Genome size	Assembly accession
Sc0000000	1475155	2302668	BraLan2	522000000	Marletaz2018
PIZI01000035v1	6646590	7457036	petMar3	1089050413	GCA_010993605.1
chr9	1668728	2337154	danRer11	1373454788	GCF_000002035.6
chr9	59367171	60565668	xenTro9	1440398454	GCF_000004195.3
chr7	15814860	16649731	galGal6	1065348650	GCF_000002315.6
chr2	73911326	75711120	mm10	2730855475	GCF_000001635.20
PIZI01000035v1	6646590	7457036	petMar3	1089050413	GCA_010993605.1

To compare sizes of clusters and of regulatory landscapes, the position of homologous genes were found using the UCSC genome browser. As it was not possible to determine the homologous gene for each species from the same data source, each source is indicated as follows: *humanProtCoding* for Aligned Ensembl Human proteins ; *AUGUSTUS AUGUSTUS* for ab initio gene predictions v3.1 ; *Genscan* for Genscan Gene Predictions ; *Mus_refSeq* for Mus Non-[Organism] RefSeq Genes. For the mouse: *Mus_refSeq* for RefSeq genes from NCBI - Annotation Release GCF_000001635.25_GRCm38.p5 (2017-08-04).

Table 2 – Mouse homologs in different chordate species.

contains the TSS coordinate ("TSS" column) for the mouse homologous gene ("Gene" column).

Assembly	Gene	TSS	source
BraLan2	Nfe2l2	1475155	humanProtCoding
BraLan2	Hoxd13	2302668	humanProtCoding
BraLan2	Hoxd9	2145400	humanProtCoding
BraLan2	Hoxd4	1873158	humanProtCoding
petMar3	Hoxd13	6646590	AUGUSTUS
petMar3	Hoxd4	7069132	AUGUSTUS
petMar3	Nfe2l2	7457036	Mus_refSeq
danRer11	Nfe2l2	1668728	Mus_refSeq
danRer11	Hoxd13	1990323	Genscan
danRer11	Hoxd9	1965727	Genscan
danRer11	Hoxd4	1951144	Genscan
danRer11	Atp5g3	2337154	Mus_refSeq
xenTro9	Hoxd4	59965982	Mus_refSeq
xenTro9	Hoxd13	59892902	Mus_refSeq
xenTro9	Nfe2l2	60565668	Mus_refSeq
xenTro9	Atp5g3	59367171	Mus_refSeq
galGal6	Hoxd4	16283021	Mus_refSeq
galGal6	Hoxd9	16298823	Mus_refSeq
galGal6	Hoxd13	16327042	Mus_refSeq
galGal6	Nfe2l2	15814860	Mus_refSeq
galGal6	Atp5g3	16649731	Mus_refSeq
mm10	Hoxd3	74578259	Mus_refSeq
mm10	Hoxd4	74721978	Mus_refSeq
mm10	Hoxd9	74697727	Mus_refSeq
mm10	Hoxd13	74668310	Mus_refSeq
mm10	Nfe2l2	75711120	Mus_refSeq

mm10	Atp5g3	73911326	Mus_refSeq

Table 3 - genomic coordinates used to plot the different loci.

qas	start	stop	name
chr6	53400000	51100000	HoxA_mm10
chr11	95250000	97550000	HoxB_mm10
chr15	101736842	104036842	HoxC_mm10
chr2	73700000	76000000	HoxD_mm10
Sc0000000	800000	3100000	Hox_BraLan2
PIZI01000035v1	5700000	8000000	Hoxdelta_petMar3
chr9	800000	3100000	HoxDa_danRer11
chr9	58800000	61100000	xenTro9
chr7	15200000	17500000	galGal6

Comparison of the four *Hox* clusters in mouse

To be consistent in measuring size of paralog regions between clusters, we used *Hox4* and *Hox13* PG, which are represented in the four clusters. *C-DOM* refers to the domain spanning from Atp5g to Hox13 paralog TSSs. *Atp5g* is absent from the *HoxA* locus, we thus ignored HoxA C-DOM. *T-DOM* refers to the domain spanning from *Hox4* to *Nfe2* paralog TSSs. *cluster* refers to the domain spanning from *Hox4* to *Hox13* paralog TSSs.

Table 4 – gene annotation at the four mouse *Hox* loci.

contains genes extracted manually from RefSeq genes from NCBI (mm10).

chr2	74668310	74671599	Hoxd13	1	+
chr2	74697727	74700208	Hoxd9	1	+
chr2	74721978	74729160	Hoxd4	1	+
chr2	73908447	73911326	Atp5g3	1	-
chr2	75704770	75711120	Nfe2l2	1	-
chr15	103018099	103036856	Hoxc4	1	+
chr15	102977032	102984444	Hoxc9	1	+
chr15	102921131	102928814	Hoxc13	1	+
chr15	103248213	103255451	Nfe2	1	-
chr11	96318267	96321638	Hoxb4	1	+
chr11	96271330	96276595	Hoxb9	1	+
chr11	96194361	96196599	Hoxb13	1	+
chr11	96072793	96075694	Atp5g1	1	-
chr11	96817414	96824127	Nfe2l1	1	-
chr6	52189687	52191703	Hoxa4	1	-
chr6	52223097	52227370	Hoxa9	1	-
chr6	52258853	52260880	Hoxa13	1	-
chr6	51432670	51458768	Nfe2l3	1	+
chr15	102662868	102671047	Atp5g2	1	-

To compensate the missing value for *HoxA* C-DOM, we used TAD coordinates obtained from GEO accession number GSE161378, downloaded under *GSE161378_E12_DL_40kb.default_domains.bed HoxA TADs* 5'-TAD is ID_0.01_2697 and 3'-TAD is (ID_0.01_2695+ID_0.01_2696) *HoxD TADs* 5'-TAD is ID_0.01_1853 and 3'-TAD is ID_0.01_1854

Table 5 – TAD sizes at the *HoxA* and the *HoxD* clusters.

HoxA_5_TAD	719999
HoxA_3_TAD	719999
HoxD_5_TAD	759999
HoxD_3_TAD	919999

SECTION II

DEVELOPMENTAL AND EVOLUTIONARY COMPARATIVE ANALYSIS OF A *HOXD* REGULATORY LANDSCAPE IN MAMMALS AND BIRDS

One of the proposed roles of TADs is to contribute to the regulation of developmental genes by forming a three-dimensional space in the nucleus that increases the probability of contact between two distant elements. That is, if distal enhancer elements must physically contact their respective target gene promoters, then TADs provide a mechanistic way for these events to happen more frequently than in their absence. The reasons for this hypothesis are that TAD boundaries frequently have deep conservation and correlate with the functional boundaries of regulatory landscapes. This hypothesis has been challenged by experiments where these relationships were tested, and there was a surprisingly weak effect on the measured changes in gene transcription. A way to reconcile these observations could be to consider that TADs would form genomic domains in which the emergence of new regulatory interactions between enhancers and promoters are favored. This increased evolvability would represent an interesting adaptive value by possibly resulting in redundancy and compensatory mechanisms.

If TAD structure assists the emergence of new regulatory elements by providing favourable conformation and appropriate transcription factors (see Delpretti et al. 2013), pre-established enhancers should promote the emergence of new elements nearby. The concept of 'regulatory hijacking', where enhancers are recruited from one developmental context to another, was proposed in (Lonfat et al. 2014). This study shows that *Hoxd13* regulations in embryonic digits and genital tubercles rely on the same topological domain, the C-DOM, involving both tissue-specific and shared enhancers. However, as external genitalia and distal limbs evolved approximately at the same period and develop at the same stage, it is hard to exclude the possibility that *Hoxd13* regulations in those tissues simply derived from an ancestral system. Moreover, the regulation of *Hox13* paralogs is more constrained than any other *Hox* genes due to severe phenotypes resulting from their dysfunction and hence, the number of evolutionary changes affecting the regulation of these genes are probably limited.

To circumvent these limitations, we analysed the regulations of anterior *Hoxd* genes in mouse somites and vibrissae placodes, which we compared to chicken somites and feather placodes. This experimental setup provides a clear chronology of tissue evolution, with somites being a chordate synapomorphy and skin appendages diversifying in amniotes. Moreover, *Hoxd1* sequence evolves fast (Guo et al. 2011) and thus, we expect to also observe changes in its regulatory modalities.

In this section, we comprehensively integrated transcription, regulatory potential, sequence conservation and chromatin architecture to gain insights in the evolution of *Hoxd1* regulation in amniotes. This work led to a first manuscript deposited on BioRxiv and is reported below.

The above work on *Hoxd1* regulation led to a collaboration with the group of Aurélien Capitan from INRAE. In this work (Allais-Bonnet et al. 2021, Annex 1), we found that the presence of supernumerary horns in goats and sheep was linked to alterations of *HOXD1* function. In polycerate goats, the region corresponding to half of the mouse T-DOM is lost. Our data demonstrated that the corresponding mouse region regulates *Hoxd1* in crest cell-derived head structures, reinforcing the likelihood that the same region in goats is involved in *HOXD1* regulation in the horn bud territory. In addition, we mapped a distinct *POLYCERATE* allele in sheep, characterized by a deletion in *HOXD1* first intron and impacting on the splicing of this gene. Our results exemplify how convergent anatomical features can arise from independent alleles altering *HOXD1* expression patterns.

DEVELOPMENTAL AND EVOLUTIONARY COMPARATIVE ANALYSIS OF A *HOXD* REGULATORY LANDSCAPE IN MAMMALS AND BIRDS

Aurélie Hintermann[1], Isabel Guerreiro[1,a], Christopher Chase Bolt[2], Sandra Gitto[1], Denis Duboule[1,2,3,#] and Leonardo Beccari[1,b]

Short title: Evolution of long-range regulation at the *HoxD* locus

Keywords: Chromatin topology, TAD, enhancers, evolution, gene regulation, development, teguments, placodes

[1]Department of Genetics and Evolution, University of Geneva, 30 quai Ernest-Ansermet, 1211, Geneva, Switzerland.
[2]School of Life Sciences, Federal School of Technology (EPFL), 1015 Lausanne, Switzerland.
[3]Collège de France, 75005 Paris, France.

[a]Present address: Oncode Institute, Hubrecht Institute-KNAW and University Medical Center Utrecht, Utrecht, The Netherlands.

[b]Present address: Institut Neuromyogène. CNRS UMR 5310, INSERM U1217, University of Lyon, Lyon, France.

[#]Corresponding author

Denis Duboule
Denis.Duboule@unige.ch

ABSTRACT

Modifications in gene regulation during development are considered to be a driving force in the evolution of organisms. Part of these changes involve rapidly evolving *cis*-regulatory elements (CREs), which interact with their target genes through higher-order 3D chromatin structures. How such 3D architectures and variations in CREs contribute to transcriptional evolvability nevertheless remains elusive. During vertebrate evolution, *Hox* genes were redeployed in different organs in a class-specific manner, while maintaining the same basic function in organizing the primary body axis. Since a large part of the relevant enhancers are located in a conserved regulatory landscape, this gene cluster represents an interesting paradigm to study the emergence of regulatory innovations. In this work, we analyzed *Hoxd* gene regulation in both murine vibrissae and chicken feather primordia, two mammalian- and avian-specific skin appendages which express different subsets of *Hoxd* genes, and compared their regulatory modalities with the regulations at work during the elongation of the posterior trunk, a mechanism highly conserved in amniotes. We show that in the former two structures, distinct subsets of *Hoxd* genes are contacted by different lineage-specific enhancers, likely as a result of using an ancestral chromatin topology as an evolutionary playground, whereas the regulations implemented in the mouse and chicken embryonic trunk rely more on conserved CREs. Nevertheless, a high proportion of these non-coding sequences active in the trunk appear to have functionally diverged between the two species, suggesting that transcriptional robustness is maintained despite a considerable divergence in CREs' sequence, an observation supported by a genome-wide comparative approach.

INTRODUCTION

Changes in the spatial and temporal regulation of genes critical for developmental processes have greatly contributed to the evolution of animal morphologies (see for example (Carroll, 2008; Rebeiz et al., 2015). The expression of such genes is generally modulated by combinations of *cis*-regulatory elements (CREs), short DNA sequences enriched for transcription factor binding sites, which tend to evolve more rapidly than the genes they control (Edwards et al., 2013; Long et al., 2016; Spitz and Furlong, 2012). CREs can be spread over large distances around the gene(s) they regulate, which gives to the spatial organization of chromatin a particular importance. Indeed, the existence of such 'regulatory landscapes' (Spitz et al., 2003) implies that enhancer-promoter interactions involve spatial proximity, which is generally thought to occur through large chromatin loops (Rao et al., 2014).

Regulatory landscapes often correspond to Topologically Associating Domains (TADs) (Andrey et al., 2013; Harmston et al., 2017), which were defined as chromatin domains where the probabilities of interactions, as measured by chromosome conformation capture (Belton et al., 2012), are higher when compared to the neighbouring regions (Dixon et al., 2016, 2012; Nora et al., 2012). The distribution of TADs tends to be conserved in various vertebrate species, a phenomenon likely associated with regulatory constraints exerted by the complex and pleiotropic regulations found around many vertebrate developmental genes. Such domains, where many interactions might occur, were proposed to form particular chromatin niches, rich in various DNA-binding proteins where the evolutionary emergence of novel regulatory sequences could be favoured due to the presence of both a pre-existing scaffold and appropriate co-factors. While genomic rearrangements altering the structure of TADs were reported to impact proper gene regulation during development leading to genetic syndromes (e.g. (Bolt et al., 2021; Lupiáñez et al., 2015), in most cases, these rearrangements also affected enhancer-promoter spatial relationships, which makes a causal assessment difficult.

Genome-wide analyses carried out in different species, comparing DNA sequence conservation, TAD organization, or some epigenetic modifications, have started to reveal some of the molecular mechanisms underlying both the evolution of gene regulation and, by extension, the modification of species-specific traits. However, integrated functional approaches are required to better understand how changes in chromatin architecture and modifications in CREs may contribute to this evolutionary process, in particular how these parameters can influence the evolvability and the robustness of gene transcription. To try to address such questions, we have used the *HoxD* locus as a paradigm of pleiotropic regulation of a multi-gene family, with a strong impact both on the control of important developmental steps, and on the emergence of evolutionary novelties.

The *HoxD* cluster contains a series of *Hoxd* genes in *cis*, which encode transcription factors that are expressed in different combinations across several embryonic structures. The cluster is flanked by two large TADs, each one matching a gene desert highly enriched in CREs (Darbellay and Duboule, 2016). These two TADs are separated by a boundary region rich in CTCF binding sites which is localized within the cluster,

close to its centromeric extremity (Andrey et al., 2013; Rodriguez-Carballo et al., 2017). The telomeric TAD comprises a large number of enhancers controlling transcription of various *Hoxd* genes in most of the tissues where HOX proteins operate (Andrey et al., 2013; Darbellay et al., 2019; Delpretti et al., 2013; Guerreiro et al., 2016; Schep et al., 2016). In contrast, the centromeric TAD contains enhancers specific for terminal structures such as digits or external genitals (Amândio et al., 2020; Montavon et al., 2011).

Amongst the many tissues where *Hoxd* genes exert a function are embryonic primordia of different dermal condensations including those of mammary glands, vibrissae and pelage hairs of mammals and those of chicken feathers (Kanzler et al., 1997, 1994; Reid and Gaunt, 2002; Reynolds et al., 1995; Schep et al., 2016)(Fig. 1). Despite the fact that these ectodermal structures share common developmental mechanisms and gene regulatory networks such as the *Wnt*, EDA and BMP signaling pathways (Biggs and Mikkola, 2014; Dhouailly et al., 2019; Di-Poï and Milinkovitch, 2016; Pantalacci et al., 2008; Wu et al., 2004), each type of primordium expresses different subsets of contiguous *Hoxd* genes, suggesting that they were co-opted in these structures following slightly different mechanisms or, alternatively, that their regulations subsequently diverged in each lineage. Since the specific topological organization of the *HoxD* locus is present in fishes (Acemel et al., 2016), it predates the emergence of hairs and feathers during amniote evolution (Dhouailly, 2009; Wu et al., 2004). For this reason, these organs represent valuable models to understand how CRE evolution within a pre-existing chromatin architecture could lead to the implementation of new gene-specific expression patterns. In contrast, the general organization of *Hoxd* gene expression along the main embryonic body axis is overall maintained across vertebrates species, in spite of the low sequence conservation observed, raising the question as to how evolving *Hox* clusters could maintain globally similar transcriptional outputs.

In this study, we try to address these issues by taking a comparative look at a well-defined regulatory landscape, positioned telomeric to the *HoxD* gene cluster, in mammals and in birds, i.e. two syntenic regions containing CREs involved either in the same regulatory tasks, or controlling regulatory aspects specific to each taxon. In the latter context, we characterized *Hoxd* gene regulation in the mouse-specific vibrissae primordia (VPs) and in the chicken-specific feather primordia (FPs). We show that different subsets of contiguous *Hoxd* genes are expressed in the mouse VP mesoderm and in the chicken FP ectoderm, indicating an independent co-option of *Hoxd* paralogs in these structures. While the transcription of these genes in VPs and FPs relies on different sets of lineage-specific CREs, located within different *Hoxd* interacting regions, these CREs exploit a largely constitutive chromatin 3D architecture.

Our comparative analysis of the regulation at work during trunk extension also revealed that *Hoxd1* displays an expression pattern markedly different from that of its neighboring paralogs, both in mouse and chick, and that this *Hoxd1*-specific expression is driven by an evolutionary conserved CRE located in a region predominantly interacting with *Hoxd1* in both species. However, the cis-regulatory code at work in the latter regulation strongly differs between the two species, illustrating the possibility that inter-species conservation of both gene transcriptional specificity and chromatin architectures are compatible with a high plasticity of the *cis*-regulatory sequences involved. The extension of this observation genome-wide suggests that conserved CRE sequences more frequently diverge in their regulatory activity rather than maintaining their functionality.

RESULTS

Divergent *Hoxd* gene regulation in mammalian and avian skin appendages.

We initially analyzed the expression profiles of *Hoxd* genes in mouse VPs and chicken FPs (Fig. 1). Skin appendage development starts with the interaction of the skin ectoderm with its underlying mesoderm, resulting in the thickening of the epithelium into a placode and in the condensation of dermal cells into a papilla (Mikkola, 2007; Sawyer and Knapp, 2003; Widelitz and Chuong, 1999). Mammalian VPs start developing at early embryonic day 12 (E12), forming a stereotyped array of eight columns and five rows of VPs (Wrenn and Wessells, 1984)(Fig. 1A). The epithelial placode subsequently invaginates and engulfs the dermal papilla forming the follicle primordium (Fig. S1). Likewise, in birds, ectoderm-mesoderm interactions result in the evagination of feather buds around embryonic day 7 (stage HH30-32 (Hamburger and Hamilton, 1951; Michon et al., 2007) (Fig. S2). Whole-mount *in situ* hybridization (WISH) revealed strong *Hoxd1* expression in the dermal papillae of the VPs (Fig. 1B; Fig. S1B). Furthermore, a comparative analysis of VPs from single E12.5 embryos revealed that *Hoxd1* transcripts accumulate slightly after ectodermal placode formation, marked by the expression of *Shh* (Chiang et al., 1999) (Fig. S1C).

Both *Hoxd3* and *Hoxd4* are also expressed in VPs, but at much lower levels than *Hoxd1* (Fig. 1B, Fig. S1E) and faint levels of *Hoxd8* mRNAs were sporadically detected in the VPs, while *Hoxd9* and more 5'-located *Hoxd* mRNAs were never observed in these structures (Fig. 1B, Fig. S1E). *In situ* hybridization on cryostat sections confirmed that *Hoxd1* transcripts accumulate in the dermal mesenchyme of VPs, but not in the ectoderm-derived portion of the follicle (Fig. S1B). Therefore, *Hoxd1-Hoxd4* are transcriptionally active in the dermal papillae of developing VPs, with *Hoxd1* being the main paralog expressed in this structure.WISH experiments in HH35 chicken embryos revealed expression of *HOXD3*, *HOXD4, HOXD8* and *HOXD9* in the FPs, while *HOXD1* and *HOXD11* mRNAs were not detected (Fig. 1C, D, Fig S2). Cryostat sections indicated that *HOXD* gene expression in chicken FPs is localized in the follicle ectoderm (Fig. S2B), in contrast with murine VPs. Different *Hoxd* genes thus operate in dermal papillae of mouse VPs and in ectodermal placodes of chicken FPs, likely reflecting the implementation of different regulatory mechanisms in the two lineages.

Hoxd gene regulation in the mouse and chicken embryonic body axis.

Hox genes are expressed in the paraxial and lateral mesoderm of the main embryonic axis, as well as in the neural tube, and display a progressively more restricted expression domain depending on their position within the cluster (Gaunt, 1988). Comparative WISH analysis of mouse and chicken embryos revealed that *Hoxd1* expression patterns in the trunk were markedly similar between the two species, yet largely divergent from that of the 5'-located neighbour *Hoxd* paralogs (Fig. 1E-G). While *Hoxd3-Hoxd8* were strongly and uniformly transcribed in the mouse paraxial and lateral mesoderm as well as in the neural tube, *Hoxd1* displayed a biphasic expression pattern. Transcripts accumulated in the embryonic tailbud, yet they rapidly disappeared from the presomitic mesoderm and remained undetected in lateral mesoderm progenitors or in the neural tube (Fig. 1F). Instead, *Hoxd1* was specifically re-activated during somite condensation, coupling the transcription of this gene to the somitogenesis clock (Dale and Pourquié, 2000; Zákány et al., 2001). Similar to the mouse pattern, the chicken *HOXD1* expression also differed from its neighboring paralogs, even though

its transcripts remained detectable in the formed somites all along the main body axis (Fig. 1G). These observations suggested that *Hoxd1* transcription in developing somites may depend on a specific set of CREs, different from CREs controlling the neighboring paralogs, and that this locus-specific regulation was evolutionarily conserved across amniotes.

Conserved paralogous regulatory landscapes.

In order to see whether these differences and similarities in regulatory specificities could be associated to variations in the global chromatin architectures at the gene cluster and the flanking telomeric TAD, we performed Capture HiC (CHiC) experiments using dissected mouse and chicken posterior trunks (Fig. 2A, B). Even though the general TAD organization at this locus is approximately two times smaller in the chicken genome, the *HOXD* region closely resembled its mouse counterpart, with the T-DOM being further organized into two sub-TADs in both species (Fig. 2, sub-TAD1 and sub-TAD2) (Rodríguez-Carballo et al., 2020; Yakushiji-Kaminatsui et al., 2018). Of note, both the presence and distribution of conserved non-coding elements (CNEs) was maintained within this similar chromatin architecture (Fig. 2, Fig. S3A).

We corroborated these observations by assessing the interactions established by the *Hoxd1*, *Hoxd4* and *Hoxd9* promoters in mouse and chicken embryonic posterior trunk cells, using 4C-seq (Fig. S3). As a reference for constitutive interactions, we used embryonic brains where *Hox* genes are not transcribed. In both species, all three viewpoints contacted sub-TAD1 with a higher probability than sub-TAD2, showing specific and non-overlapping regions where interactions were enriched. *Hoxd1* contacted the first half of sub-TAD1 (Fig. S3, D1) more intensively than *Hoxd4* or *Hoxd9.* In turn, *Hoxd4* contacts within the 3' half of *HoxD* sub-TAD1 were increased up to the sub-TAD boundary (Fig. S3, D4). Finally, *Hoxd9* interactions were mostly limited to the sub-TAD boundary region (Fig. S3, D9), where they were higher than for *Hoxd4*. These preferentially interacting regions within the 3' TAD were found in both brain and posterior trunk cells, yet with an overall lower contact frequency in the brain, reflecting the default and transcription-independent conformation of the region, as shown in (Andrey et al., 2013; Dekker and Heard, 2015; Noordermeer et al., 2011).

Chromatin architectures largely rely upon the zinc-finger protein CTCF which, together with the cohesin complex, can induce the formation and stabilisation of large loops. (Hansen et al., 2017; Nanni et al., 2020; Pugacheva et al., 2020). The ChIP-seq analysis of the CTCF profiles revealed that both at the mouse and chicken *HoxD* loci, a bound CTCF site is observed between the D1 and D4 regions (Fig. S3, blue arrowhead). Also, as previously reported (Rodriguez-Carballo et al., 2017; Rodríguez-Carballo et al., 2020; Yakushiji-Kaminatsui et al., 2018), a cluster of CTCF binding sites delimits the boundary between the mouse and the chicken sub-TAD1 and sub-TAD2, thus making the transition between the D4 and D9 regions in both species (Fig. 2; Fig. S3). These results suggest that the evolutionarily conserved contact topology of 3' *Hoxd* genes within the T-DOM may contribute to the specific expression of *Hoxd1* in mouse and chicken embryonic structures.

Searching for similar and divergent inter-species regulatory sequences.

To identify potential regulatory sequences within those regions showing preferential interactions with *Hoxd* target genes in either the VPs, the FPs and in posterior trunk cells, we carried out ChIP-seq analyses

using an antibody against H3K27ac, a histone modification indicative of an active chromatin state (Creyghton et al., 2010; Rada-Iglesias et al., 2011). The profile of H3K27ac in fetal brains was used as a negative control to identify putative tissue-specific CREs after microdissections of a comparable piece of fetal skin containing the primordia of either VPs or FPs. In all H3K27ac profiles, the T-DOM contained most of the tissue-specific enrichment, whereas H3K27ac signals over the C-DOM were virtually absent (Fig. 3).

We identified those genomic regions residing outside of genes and promoters and accounting for the highest tissue-specific H3K27ac enrichment (hereafter referred to as Most Acetylated Regions, MARs; see materials and methods). We annotated the MARs of both the mouse VPs and the chicken FPs (Fig. 3, red rectangles) and compared them to the extent of the regions preferentially interacting with *Hoxd1*, *Hoxd4* and *Hoxd9* promoters (D1, D4 and D9 respectively)(Fig. 3). In the mouse VPs, the MARs mostly mapped within the D1 region (8 tissue-specific H3K27ac -positive elements)(Fig. 3A; Fig. S4A), in agreement with the relatively stronger expression of *Hoxd1* there, when compared to its neighbors (Fig. 1B; Fig. S1D). Instead, in chicken dorsal skin samples, where *HOXD1* is not active (Fig. 1D), only two H3K27ac-positive sequences were located within the D1 region and the MARs largely overlapped the D4-D9 interacting region (Fig. 3A, red rectangle; Fig. S4A). Therefore, VPs and FPs display different distributions of H3K27ac-positive elements across the T-DOM, in agreement with the specific *Hoxd* gene subsets that are expressed in these tissues. In contrast, the MARs of mouse and chicken PT samples both overlapped the D4 region (Fig. 3B, red rectangles; Fig. S4B). This result is in agreement with the similar expression patterns in the embryonic trunk of the two species, and with the domain of expression of *Hoxd4* being more extended than that of *Hoxd9*. Of note, different H3K27ac-positive elements were identified within the mouse and chicken D1 region, suggesting that they could contribute to the specific transcription of *Hoxd1* in the developing somites.

Eutherian-specific elements controlling *Hoxd1* in the vibrissae progenitors.

We further characterized the interactions between the *Hoxd1, Hoxd4* and *Hoxd9* promoters with the D1 region by performing 4C-seq analyses in microdissected E12.5 VPs (Fig. S5A, orange profiles), using brain cells as a negative control (Fig. S5A, grey profiles, overlaps in braun). We observed that, while the interactions established by the active *Hoxd1* gene increased specifically within region D1 in the VP sample when compared to the brain, *Hoxd4* displayed only a modest augmentation of contacts in the same region and the interactions with *Hoxd9* remained essentially the same in both tissues (Fig. S5A), as expected from these latter two genes, which are not transcribed in the VPs. In contrast, when the profiles obtained from posterior trunk cells, where both *Hoxd4* and *Hoxd9* are active, were subtracted from the VP profiles, a gain of both *Hoxd4* and *Hoxd9* reads were observed in their respective binding D4 and D9 regions (Fig. S5A, dark braun profiles inbetween the 4C profiles). Indeed, gained signals for the *Hoxd4* viewpoint were mostly scored over the D4 region whereas the D9 region was enriched in contacts established by *Hoxd9* (Fig. S5A).

We then analyzed *Hoxd1* expression in the VPs of mouse embryos carrying different deletions of the T-DOM (Fig. 4A, B). In the $HoxD^{Del(AttP\text{-}TpSB2)lac}$ allele, a 150kb large piece of DNA directly adjacent to the *HoxD* cluster was removed (hereafter referred to as Del(Attp-SB2). Since Del(Attp-SB2) homozygous embryos are not viable due to the deletion of the essential *Mtx2* gene (Andrey et al., 2013), we crossed this line with the mouse line $HoxD^{Del(1\text{-}13)d9Lac}$ (Spitz et al., 2001) hereafter referred to as Del(1-13). In this line, the

entire *HoxD* cluster is replaced by a *Hoxd9::LacZ* transgenic cassette (Fig. 4C), such as to be able to see the effect of the $HoxD^{Del(AttP\text{-}SB2)lac}$ deletion upon *Hoxd* gene regulation *in cis*. As expected, *Hoxd1* and *Hoxd4* expression in the VPs of trans-heterozygous $HoxD^{Del(AttP\text{-}TpSB2)Lac/Del(1\text{-}13)}$ mutant embryos was completely abolished, in contrast with Del(1-13) heterozygous control littermates. Conversely, the $HoxD^{Del(TpSB2\text{-}TpSB3)lac}$ allele (hereafter Del(SB2-SB3)(Andrey et al., 2013) carries a large deletion spanning the second half of sub-TAD1 and the entire sub-TAD2. We detected robust *Hoxd1* expression in the VPs of E12.5 Del(SB2-SB3) homozygous mutant embryos (Fig. 4A).

We complemented these results by producing mouse transgenic embryos carrying bacterial artificial chromosomes (BACs) spanning the *HoxD* cluster and/or parts of sub-TAD1 (Fig. 4C). In agreement with the analysis of the T-DOM deletions, the BACs covering either the *HoxD* cluster (BAC^{HoxD} (Schep et al., 2016) or mapping outside of the Attp-SB2 region (BAC^{T1} (Delpretti et al., 2013) did not display any LacZ reporter activity in E12.5 to E13.5 VPs. Instead, mouse embryos transgenic for BAC^{Mtx2} (Allais-Bonnet et al., 2021), which covers most of the Attp-SB2 region (and corresponds closely with the D1 region), displayed strong reporter expression in the VPs (Fig. 4B). These results demonstrate that the Attp-SB2 region is both necessary and sufficient to drive *Hoxd* gene expression in the developing VPs.

Together with the WISH and 4C-seq data, these results delimited a segment of 148kb hereafter referred to as the VP regulatory region (Fig. 4C, VPRR), a region that closely overlaps with VP-specific MAR region (Fig. 3A). In particular, the comparison between the H3K27ac profiles in E12.5 VPs and brain cells identified at least four non-coding sequences (or groups thereof-) located within the VPRR and referred to as VP acetylated elements (VPAEs)(Fig. 4D, open arrowheads). Phylogenetic footprinting analysis revealed that all four elements are conserved across Therians, which include placental mammals and marsupials (Fig. S5B). Noteworthy, a core VPAE3 sequence is also conserved in reptiles and birds, suggesting that this sequence evolved earlier in the tetrapod lineage. To assess the regulatory potential of these sequences, we cloned small regions containing the four VPAEs into a *LacZ* reporter vector upstream of the *ß-globin* minimal promoter and tested them in transient transgenic assays (Fig. 4D, tgVPAE1-4; Fig. S5C). Amongst the tested CREs, VPAE2 displayed strong and reproducible enhancer activity in VPs (Fig. 4E), comparable to that observed in embryos transgenic for BAC^{Mtx2}. Instead, ß-galactosidase activity in VPs was not scored for VPAE1, VPAE3, or VPAE4 transgenic embryos (Fig. S5C). These results indicated that the expression of *Hoxd1* to *Hoxd4* in VPs depends at least on the VPAE2 enhancer, located 120 kb 3' to the *Hoxd1* transcription start site within the D1 region. Noteworthy, in Del(Attp-SB2) mouse embryos, expression of *Hoxd1* was also abolished from facial mesenchymal progenitors, suggesting that this regulatory region includes other CREs controlling this gene.

Lineage-specific regulatory sequences in chicken feather buds.

We also examined the evolutionary conservation of non-coding elements specifically enriched in H3K27ac marks in micro-dissected dorsal skin samples of chick embryos and located within the MARs-containing region defined for this structure (Fig. 3A). While the vast majority of these elements could not be identified in mammals (Fig. S6), they were found conserved across birds and reptiles indicating a lineage-specific evolution of CREs driving *Hoxd* gene expression in FPs, analogous to what was observed for the mouse VPs. These results further support the idea that the evolutionary emergence of lineage-specific CREs

within an ancestral chromatin topology may have triggered differential *Hoxd* gene regulation in mammalian and avian skin appendages.

Tetrapod-conserved regulatory elements active in forming somites.

The expression of *Hoxd1* during the extension of the mouse and chicken embryonic axes differs substantially from that of the neighboring *Hoxd* genes (Fig. 1B, D; Fig. S1E). We could thus expect that the accompanying regulations might involve *Hoxd1*-specific interacting elements, distinct from the MARs identified for the mouse and chicken embryonic PTs (Figs. 3B and S4B). To try to localize the CRE(s) active over *Hoxd1* in forming somites, we initially looked at different alleles where deletions had been engineered (Fig. 4F, G). *Hoxd1* expression was abolished in developing somites of Del(Attp-SB2) E9.5 embryos, compared to their control Del(1-13) littermates (Fig. 4F), while similar transcript levels were maintained in the tailbuds, suggesting that *Hoxd1* regulation in both structures involves different regulatory sequences. Accordingly, embryos transgenic for BAC^{Mtx2} displayed strong LacZ staining in somites but not in the tailbud, with stronger activity in the last formed somites, thus closely mimicking *Hoxd1* transcription in these structures (Zákány et al., 2001). In contrast, *LacZ* expression was uniform along the paraxial mesoderm of TgBAC^{HoxD} transgenic embryos, yet without showing any *Hoxd1*-like specific reporter activity in the forming somites (Fig. 4G).

The H3K27ac profiles revealed that most of the signals specific for posterior trunk cells (SAEs) were located in the D4 sub-region, a region involved in *Hoxd* gene regulation in paraxial and lateral plate mesoderm derivatives (Fig. 4G, tgBAC^{T1}). Amongst those H3K27ac peaks mapping within the genomic interval spanned by the Del(Attp-SB2) mutation and covered by BAC^{Mtx2}, two were assessed by transgenic analysis (Fig. 4H; black arrowheads). While the SAE sequence within VPAE3, conserved throughout tetrapods, gave expression in the forming somites (Fig. 4I), the other SAE sequence, which was not conserved except in Eutherians, did not show any signal in these structures (Fig. S5B, D). The chicken sequence orthologous to mouse VPAE3 was also found enriched in H3K27ac in chicken posterior trunk cells (Fig. S4B), suggesting that it is functionally conserved in mammals and birds. These results indicated that the regulation of *Hoxd1* in the forming somites depends on a specific set of enhancers, which are in part conserved across amniotes and differ from those controlling other *Hoxd* genes in the paraxial and lateral mesoderm.

Conservation and divergence of *HoxD* enhancers in mouse and chicken embryonic trunk cells.

When mouse VP and dorsal chicken skin were used, the H3K27ac profiles revealed that the vast majority of H3K27ac-positive elements are not conserved between the two species. This is in contrast with posterior embryonic trunk cells, where the proportion of H3K27ac-positive elements coinciding with a conserved sequence was considerably higher. Surprisingly, however, several conserved elements enriched in H3K27ac in either the mouse or the chicken posterior trunk cells, were not enriched in the same tissue of the other species (Fig. S4), illustrating a divergence in the combinations of CREs at work in this dissected piece of embryonic tissue, in contrast with the otherwise well-conserved chromatin architectures and expression patterns.

We next wanted to assess whether this discrepancy between the inter-species sequence conservation on the one hand, and their H3K27ac coverage in the same developing tissues on the other hand, would also

occur outside the *HoxD* locus. We compared H3K27ac profiles across different tissues (proximal and distal forelimb buds, fetal skin, brain and posterior trunks) of mouse and chicken embryos genome-wide and related them to sequence conservation. In both mouse and chicken, the majority of non-coding H3K27ac-positive elements mapping more than 2kb away from a gene transcription start site (putative enhancers, hereafter referred to as enhancers) displayed H3K27ac enrichment in only one or the other of the embryonic structures analyzed (Fig. 5A). Instead, most H3K27ac-positive elements located less than 2kb far from the transcription start sites ('promoters') were enriched across different samples (Fig. 5A), a result in agreement with the generally pleiotropic expression of genes with important developmental functions. However, the proportion of enhancers displaying H3K27ac in more than one tissue was significantly increased when considering only non-coding sequences which were evolutionary conserved between mouse and chicken (Fig. 5A; p-value < e-15, Fisher test).

In agreement with what was observed at the *HoxD* locus, the majority of enhancers evolutionarily conserved between mouse and chicken differed in their H3K27ac enrichment patterns between the two species (Fig. 5B, C). In fact, 79 percent (1141/1445) of the enhancers labelled by H3K27ac in only one of the embryonic tissues analyzed in chicken were either not active in any of the corresponding mouse samples, or active in different mouse embryonic structures. Accordingly, H3K27ac-positive enhancers scored in different chicken samples frequently diverged in their H3K27ac coverage when compared to the related mouse tissues (Fig. 5B). This data indicates that DNA sequence similarity between two species can be maintained in regulatory sequences despite divergences in their tissue-specific functionalities.

DISCUSSION

Versatile *Hoxd* gene functions in skin appendages.

Hox genes play a dual role in skin appendage development. While their colinear expression is initially necessary to pattern the skin dermis (Kanzler et al., 1997), they are subsequently transcribed in the dermal papillae and/or ectoderm of embryonic and adult hair pelage follicles and vibrissae precursors (Godwin and Capecchi, 1998; Packer et al., 2000; Reynolds et al., 1995; Yu et al., 2018), as well as in the avian buds (Chuong et al., 1990; Kanzler et al., 1997). Here we show that the dermal papillae of embryonic mouse vibrissae express *Hoxd1* to *Hoxd4*, with the former gene's mRNAs being the first detected, followed by considerably lower levels of *Hoxd3* and *Hoxd4* RNAs. The transcription in VPs is driven by CREs located in a region encompassing approximately 120kb of DNA adjacent to the telomeric end of the mouse *HoxD* cluster, a region that is preferentially contacted by the *Hoxd1* gene. Therefore, a proximity effect is observed in the building of a general chromatin architecture, based on constitutive contacts between the *Hoxd1* region and the neighbouring sub-TAD, whereby most of the *Hoxd1* regulatory sequences are located in the adjacent portion of the regulatory landscape.

While a detailed functional characterization of *Hoxd* genes during the development of whiskers was not the aim of this work, we did not observe any major morphological alteration in the vibrissae of mice carrying genetic deletions of this precise region, despite the complete abrogation in the expression of *Hoxd1* to *Hoxd4* in these structures. This lack of any visible abnormal phenotype may be due to functional redundancy

with other *Hox* paralogy groups expressed there, such as *Hoxc8* (Yu et al., 2018), to a lack of analytical power or of the appropriate behavioural paradigms that may have revealed abnormally function whiskers. The diverse expression of *Hox* paralogs in mammalian and avian skin appendages nonetheless illustrates the plasticity in the usage of *Hox* gene functions in these structures. This plasticity likely relies upon the independent acquisition of specific CREs for each gene cluster, as these skin derivatives emerged after the genome duplication events thought to be at the origin of the vertebrate lineage (Holland et al., 2008; Holland and Garcia-Fernàndez, 1996). This is also also supported by the observation that all VPAEs and FPAEs identified in this study are mammalian or avian specific.

While *Hoxd9, Hoxd11* and *Hoxd13* are expressed in the ectoderm-derived hair fibers (Godwin and Capecchi, 1998; Packer et al., 2000; Reynolds et al., 1995; Yu et al., 2018), we did not observe any *Hoxd1* in the vibrissae shaft, nor any LacZ activity when the TgBAC^{Mtx2} (carrying the VPRR) transgenics were used, indicating that these expression patterns may depend on different sets of CREs. Our data also show that the expression of *Hoxd3* to *Hoxd9* in the ectoderm of the chicken FPs depends on CREs unrelated to those operating in the mouse VPs. These observations may indicate that different sets of *Hoxd* genes were independently co-opted during the evolution of skin derivatives, likely through *de novo* enhancer acquisition in the mammalian and avian lineages, in a way related to the regulation of *Hoxd9 and Hoxc13* in the embryonic mammary buds and in the nails and hairs of mammals, respectively (Fernandez-Guerrero et al., 2020; Schep et al., 2016). Alternatively, the rapid pace of tegument evolution in tetrapods may have been accompanied by a divergence in enhancer sequences, which makes comparisons difficult. In this context, we can not rule out the possibility that *Hoxd* gene regulation in skin appendages of both mammals and birds derives from an initial pan-*Hoxd* expression pattern both in the ectoderm and mesoderm of skin appendages, which then evolved differently in each lineage through modification of the CRE complement located within the adjacent TAD. The fact that two CNEs acetylated in mouse VPs and chick FPs are also acetylated in mouse and chick posterior trunk cells may indicate that the former structures co-opted parts of the regulatory mechanisms at work in the latter. In this view, DNA sequence conservation may be due to constraints imposed by the function in the trunk rather than by the function in the skin.

Enhancer acquisition in evolving tetrapod skin appendages.

TADs which contain pleiotropic genes of key importance for development are frequently conserved across tetrapods and topological changes in their organization often correlate with important modifications in gene expression (Eres et al., 2019; Liao et al., 2021; Torosin et al., 2020; Yakushiji-Kaminatsui et al., 2018). Our results show that despite a considerable divergence in the non-coding elements localized in the mouse and chicken T-DOM, the global pattern of interactions is conserved leading to similar chromatin topologies. Indeed, in both cases, different *Hoxd* genes interact preferentially with distinct portions of sub-TAD1. These contacts are for the most constitutive, even though we observed an activity-dependent increase in interactions in the mouse VPs, and in mouse and chicken PT cells. Of interest, the regions within T-DOM, which were preferentially contacted by the *Hoxd1*, *Hoxd4* or *Hoxd9* genes (D1 to D9) are organized in a linear sequence opposite to the order of the genes themselves, as if *Hoxd1* was folding on its closest possible chromatin environment within T-DOM, bringing along *Hoxd3* towards the next chromatin segment and positioning the

interactions involving *Hoxd9* further away into T-DOM. This sequential positioning of interactions in mirror with spatial colinearity may impose a generic contact pattern, whereby *Hoxd9*, for example, is unable to establish strong interactions with the D1 region and its enhancers, for this region is primarily contacting *Hoxd1*, thus imposing constraints upon the distributions of interactions (see below).

Furthermore, the distribution of H3K27ac-positive regions in the chicken FPs and mouse VPs across these interacting regions correlated well with the specific subsets of *Hoxd* genes expressed in each of these tissues. Finally, the regulation of *Hoxd1* in developing somites relies on evolutionary conserved elements located in the region D1. This general regulatory topology is close to that observed in teleost fishes (Acemel et al., 2016; Woltering et al., 2014), suggesting that it was already in place at the root of the vertebrate lineage and that its hijacking by nascent enhancer sequences, may have favored the co-option of specific *Hoxd* gene subsets as targets of the novel regulations. The regulation of *Hoxd* genes in the mouse VPs and chicken FPs, as well in the mouse mammary buds (Schep et al., 2016) may illustrate this process.

Plasticity of the *cis*-regulatory code *versus* conservation in TAD organization.

Due to their functional relevance in the regulation of gene expression, CREs tend to be evolutionarily conserved across species (Sandelin et al., 2004; Sanges et al., 2013, 2006). However, several studies have shown that enhancers controlling developmental gene expression can be functionally conserved even across distant phylogenetic lineages despite a strong divergence in their DNA sequences (Ambrosino et al., 2019; Berthelot et al., 2018; Wong et al., 2020) (Eichenlaub and Ettwiller, 2011; Schmidt et al., 2010; Villar et al., 2015). Our inter-species comparison of CNEs enriched in H3K27ac across different embryonic structures revealed that a considerable portion of evolutionary conserved CREs diverged in their patterns of activity between the two species analyzed, in agreement with previous observations (Dermitzakis and Clark, 2002; Schmidt et al., 2010; Vierstra et al., 2014; Villar et al., 2015). This dichotomy between the conservation of enhancer sequences and their function is illustrated by the CREs controlling *Hoxd1* expression in the mouse and chicken developing somites. While some of the sequences active in the mouse posterior trunk are evolutionary and functionally conserved in chicken, such as VPAE3, which can drive LacZ expression in the last formed somites of the murine embryo, other elements conserved at the nucleotide level displayed H3K27ac enrichment only in the mouse.

In fact, a large fraction of H3K27ac-positive sequences detected either in the mouse or in the chicken PT cells were not conserved between the two species, suggesting that they evolved in a species-specific manner. In contrast, the high proportion of chicken promoters that display similar H3K27ac enrichments in mice, together with the similar expression patterns observed for mouse and chicken *Hoxd* genes in embryonic trunk cells, demonstrate that conserved gene expression can occur despite a high evolutionary plasticity in the *cis*-regulatory code usage, in agreement with previous reports (Berthelot et al., 2018; Snetkova et al., 2021). This may point to a regulatory strategy where multiple enhancers act coordinately to regulate the same target gene(s) thus creating a quantitive effect in a way similar to the reported action of super-enhancers (Sabari et al., 2018), yet where various subsets of CREs may display distinct tissue-specificities. This process may be implemented preferentially within large regulatory landscapes, as in the case of *Fgf8* (Marinic et al., 2013) or the TAD controlling *Hoxd* gene transcription in external genitals (Amândio et al., 2020).

A regulatory playground with constraints.

In those cases that were investigated, the respective positions of regulatory sequences within large syntenic regulatory landscapes are generally conserved, and the global regulatory architecture found at one particular developmental locus in one species can be extrapolated to its cognate locus in another amniote species (e.g. (Kragesteen et al., 2018), implying that enhancer-promoter interactions are also evolutionary well conserved, even for CREs acting over long distances (Irimia et al., 2012; Laverré et al., 2021), although some mechanistic elements may vary (Ushiki et al., 2021). In agreement with this, and contrasting with the apparent divergence in the *cis*-regulatory code controlling *Hoxd* gene expression in mouse and chicken embryos, our phylogenetic footprinting and epigenetic analysis revealed that non-coding elements conserved in their DNA sequence between the mouse and chicken genomes were distributed across the two species in a very similar manner. Therefore, the potential gain and/or loss in the two lineages of particular CREs did not significantly impact the internal organization of those conserved regulatory elements, nor did they impact upon the global chromatin topology of the locus. This confirms the observation that even the significant divergence in CREs located at the *HoxD* locus, which accompanied the evolution of the snake body plan, did not result in any substantial alteration of the overall corn-snake *Hoxd* gene interaction map (Guerreiro et al., 2016). Altogether, these observations suggest that internal organization of the TADs at the *HoxD* locus and the specific interaction profiles for various subsets of *Hoxd* target genes are resilient to considerable variation in their CREs.

These results also highlight the somewhat dual properties that large regulatory landscapes may have imposed in the course of regulatory evolution. On the one hand, a pre-established chromatin topology, as materialized by TADs, may have provided unique 'structural niches' to evolve new enhancer sequences, due to the proximity of factors already at work. This might especially apply to landscapes where a strong quantitative parameter may be instrumental, and *a fortiori* when the target gene(s) are acting in combination, coding for proteins of rather low specificities (Bolt and Duboule, 2020). This may be illustrated nowadays by large landscapes, where all kinds of enhancer sequences are mixed within the same TAD. On the other hand, this regulatory playground, while stimulating the emergence of regulations, will constrain their realms of action due both to the very architecture that favours their evolution and to the accessibility to 'a' subset of contiguous target genes as a result of the global distribution of interactions at this precise place of the landscape.

MATERIAL AND METHODS

Mouse strains and chicken eggs

All mutant strains are listed in Table S1 and were backcrossed continuously to Bl6/CBA mixed animals and maintained as heterozygous stocks. Chick embryos from a White Leghorn strain were incubated at 37.5°C and staged according to Hamburger and Hamilton (1951). For genomics analyses, we used the mouse assembly mm10 and the chicken galGal6.

Enhancer cloning and transgenics

The different CRE candidate sequences were amplified using PCR-specific primers (see Table S2) and the Expand High Fidelity PCR system (Roche). Amplified fragments were gel-purified and cloned into the pSK-LacZ vector as in (Beccari et al., 2016). For the transgenesis assays, each LacZ reporter construct was digested

with NotI and KpnI and the restriction fragment encoding the potential enhancer sequence, the beta-globin minimal promoter and the LacZ reporter gene was gel-purified and injected into the male pronucleus of fertilized oocytes. F0 embryos were dissected either at E10, or at E12, fixed and stained for LacZ activity according to standard protocols.

In situ hybridization

The probes used in this study for *in situ* hybridization were either previously reported, or produced by PCR amplification of a fragment subsequently ligated in pGEMT Easy vector (Promega). Primers used to amplify the fragment of interest are listed in Table S2. Digoxigenin (Dig) labeled probes were synthesized *in vitro* by linearizing the respective plasmids using specific restriction enzymes and transcribed *in vitro* either with T7/T3, or Sp6 RNA polymerase (Table S3) and the Dig-RNA labeling mix (Roche). The probes were purified using the RNA easy mini kit. WISH experiments were performed as described in (Woltering et al., 2009). Probes are listed in (Table S3). *In situ* experiments in sections were performed according to (Sockanathan and Sockanathan, 2015). Cryostat sections were prepared after fixing embryos in 4% PFA solution at 4ºC overnight. Subsequently, embryos were washed three times with PBS and passed to increasing sucrose solutions at 4ºC in agitation (15% sucrose solution for 3-4 hours, followed by incubation in 30% sucrose solution overnight)and finally included in OCT blocks. Sections of 20um were obtained with a Leica cryostat and stored at -80ºC.

ChIP-seq

ChIP-seq experiments were performed using the ChIP-IT High sensitivity kit (active motif according to manufacturer instructions with minor modifications). For the mouse samples, 12 pairs of E12 VPss, or one hundred E8.75. PT regions were used, the latter dissected at the level of the 2nd to 4th pair of somites. For the chicken samples, the dorsal skin of five HH35 embryos or the PT or ca. 80 HH19-21 PT regions were isolated. In all cases, the samples were fixed for 10 min in 1% formaldehyde at RT and the crosslinking reaction was quenched with Glycine. Subsequently, nuclei were extracted and sonicated in (Tris HCl pH 8.0 50mM, EDTA 10mM, 1% SDS) to an average fragment size of 200-500bp using a Diagenonode Bioruptor Sonicator. Ca. 10 to 20 µg of sonicated chromatin were diluted tenfold with Dilution buffer (HEPES pH 7.3 20mM, EDTA 1mM, NaCl 150mM, NP40 0.1%) and incubated with 2 µg of anti-H3K27ac antibody (Abcam, ab4729). All buffers contained 1X Complete Proteinase Inhibitor cocktail (EDTA-free; Roche) and 10mM Sodium Butyrate to prevent protein degradation and Histone de-acetylation. Chromatin–antibody complexes were immunoprecipitated with Protein G-coupled agarose beads-purified and de-crosslinked according to manufacturer instructions. Approx 5 to 10 ng of purified DNA were used for ChIP library preparation by the Geneva IGE3 Genomics Platform (University of Geneva) and sequenced to obtain 100-bp single-end reads on an Illumina HiSeq2500 or HiSeq4000 system. ChIP-seq reads processing was done as in (Amândio et al., 2020), using the laboratory Galaxy server (Afgan et al., 2016). Adapters and bad-quality bases were removed with Cutadapt version 1.16 (Martin, 2011) (options -m 15 -q 30 -a GATCGGAAGAGCACACGTCTGAACTCCAGTCAC). Reads were mapped to the mouse genome (mm10) and to the chicken genome (galGla6) using Bowtie2 (v2.3.4.1) (Langmead and Salzberg, 2012)

with standard settings. The coverage was obtained as the output of MACS2 (v2.1.1.20160309) (Zhang et al., 2008) Peak calling was done using MACS2 (v2.1.0.20160309). CTCF motif orientation was assessed using the CTCFBSDB 2.0 database (Ziebarth et al., 2012) with EMBL_M1 identified motifs. H3K27ac ChIP-seq were normalized with the MAnorm model (Shao et al., 2012) using peaks called with MACS2 and a fixed extension of 200 bp. MARs were annotated within the mouse T-DOM (chr2:74,872,605-75,696,338) and the chicken T-DOM (chr7:15856252-16252401). CNEs were used to define corresponding regions in each species. The brain coverage was subtracted from each input file. Each possible genomic window totalizing 30 % (VP) or 50% (PT) of the total T-DOM H3K27ac coverage was computed and the smallest and densest window was annotated as 'MAR'.

4C-seq experiments

4C-seq experiments were performed according to (Noordermeer et al., 2011). Briefly, the tissues were dissected, dissociated with collagenase (Sigma Aldrich/Fluka) and filtered through a 35 micron mesh. Cells were fixed with 2% formaldehyde (in PBS/10%FBS) for 10 min at RT and the reaction was quenched on ice with Glycine. Nuclei were extracted using a cell lysis buffer (Tris pH 7.5 10 mM, NaCl 10 mM, $MgCl_2$ 5 mM, EDTA 0.1 mM,) and stored at -80°C. Approximately 20 pairs of mouse E12 VPs, or 200 E9.5 PT regions dissected at the level of the 2^{nd} to 4^{th} pair of somites, as well as 120 chicken HH19-21 PT samples, were used. Nuclei were digested with NlaIII (New England Biolabs) and ligated with T4 DNA ligase HC (Promega) in diluted conditions to promote intramolecular ligation. Samples were digested again with DpnII (New England Biolabs) and re-ligated with T4 DNA ligase HC. These templates were amplified using Expand long template (Roche) and PCR barcoded primers flanked with adaptors. For each library, 8 to 10 independent PCR reactions were pooled together and purified using the PCR purification kit (Qiagen). Multiplexed libraries were sequenced on Illumina HiSeq 2500 to obtain 100 bp single-end reads. Demultiplexing, mapping and 4C-seq analysis were performed using a local version of the pipeline described in (David et al., 2014), on the mouse GRCm38 (mm10) and chicken GalGal6 genome assemblies. The profiles were smoothened using a window size of 11 fragments and normalized to the mean score in +/-5 Mb around the viewpoint. When multiple independent biological replicates were available, average 4C-seq profiles were calculated.

Capture HiC-seq experiments

Libraries of sureSelect enrichment probes were designed as described in (Bolt et al., 2021) for mouse and (Yakushiji-Kaminatsui et al., 2018) for chicken. Dissected tissues were processed as in (Yakushiji-Kaminatsui et al., 2018) with the following changes: cells were cross-linked in 2% formaldehyde/PBS and washed three times instead of being quenched by glycine. HiC-libraries were prepared as in (Yakushiji-Kaminatsui et al., 2018).

ETHICS APPROVAL

All experiments involving animals were performed in agreement with the Swiss Law on Animal Protection (LPA), under license No GE 81/14 (to DD).

DATA AVAILABILITY

All raw and processed datasets are available in the Gene Expression Omnibus (GEO) repository under accession number GEOXXXXXX

COMPETING INTERESTS

The authors declare that they have no competing interests.

FUNDING

This work was supported by funds from the University of Geneva, the Ecole Polytechnique Fédérale (EPFL, Lausanne), the Swiss National Research Fund (No. 310030B_138662) and the European Research Council grants System*Hox* (No 232790) and Regul*Hox* (No 588029) (to D.D.). Funding bodies had no role in the design of the study and collection, analysis and interpretation of data and in writing the manuscript.

AUTHORS CONTRIBUTIONS

A.I. : Designed experiments, realized experiments and interpreted the results, wrote the manuscript

I.G. : Did some cloning and *in situ* hybridizations

C.C.B. : Provided unpublished datasets and helped for some the C-HiC experiments

S.G. : Genotyped and helped analyze mouse mutants.

D.D. : Designed experiments, interpreted some results and wrote the paper.

L.B. : Designed experiments, interpreted some results and wrote the paper.

ACKNOWLEDGEMENTS

We thank Dr Lucille Lopez-Delisle for her help with bioinformatics as well as members of the Duboule laboratories for comments and discussions and Gregory Loichot for artwork. This work was supported in part by using the resources and services of both the Gene Expression Core Facility at the School of Life Sciences of the EPFL in Lausanne and the genomic plateform from the IGE3 centre of the University of Geneva. Transgene injections were performed at the transgenesis platform of the University of Geneva Medical School. This work was supported by funds from the University of Geneva, the Ecole Polytechnique Fédérale (EPFL, Lausanne), the Swiss National Research Fund (No. 310030B_138662) and the European Research Council

grants SystemHox (No 232790) and RegulHox (No 588029) (to D.D.). Funding bodies had no role in the design of the study and collection, analysis and interpretation of data and in writing the manuscript.

LEGENDS TO FIGURES

Figure 1. Transcription of *Hoxd* genes in mammalian and avian skin primordia. A. Schematic illustration of a E12.5 mouse embryo with VPs (Vibrissae Primordia) represented as red circles. The dorso-ventral (D, V) and the caudo-rostral (C, R) axes are shown as red arrows representing the timing of appearance of the vibrissae primordia. **B.** WISH on E12.5 faces. The arrowheads point to the most caudo-dorsal placode, which is the largest one and the first to appear. Staining intensity progressively decreases from *Hoxd1* to *Hoxd8.* **C.** Schematic illustration of the back of a HH35 chicken embryo with feather placodes as red circles. A, anterior; P, posterior. **D.** WISH on HH35 chicken showing the skin of the upper back. The arrowheads point to the neck at the level of the shoulders. *HOXD1* is not expressed, whereas *HOXD3* and *HOXD8* signal intensities are stronger than for *HOXD4* and *HOXD9.* **E.** Schematic representation of a E9.5 or HH20 posterior trunk and tailbud. A, anterior; P, posterior; PSM, presomitic mesoderm; NT, neural tube; IM, intermediate mesoderm; FS, forming somites; PM, paraxial mesoderm; LM, lateral mesoderm; S, somites. **F.** WISH on E9.5 mouse posterior trunk and tail bud. *Hoxd1* transcription pattern is different from that of *Hoxd3* and *Hoxd4.* **G.** WISH on HH20 chicken posterior trunk and tail bud. *HOXD1* transcription pattern is different from *HOXD3* and *HOXD4.* The white arrowheads in panels **F** and **G** point to the *Hoxd1* stripe in the PSM. The black arrowheads show the difference between mouse (**F**) and chicken (**G**) in the *Hoxd1* mRNAs persistence in formed somites.

Figure 2. Capture HiC-seq at the mouse and chicken *HoxD* loci. A, **B.** Capture HiC-seq heatmaps using dissected posterior trunk cells. Below each heatmap, a CTCF ChIP-seq track is shown, produced from the same material. The similarities in the structural organization of the mouse and the chicken loci are underlined either by the positions of syntenic CNEs at key positions (numbered vertical dashed lines), the domains produced by TAD calling (black bars below) and the presence of a sub-TAD boundary (asterisk) within T-DOM. Genes are represented by empty rectangles. In panel **A,** E9.5 mouse PT cells were used and the CHiC heatmap was mapped on mm10 with a bin size of 5 kb (chr2:73800000-75800000). In panel **B,** HH20 chicken PT cells were used and the CHiC heatmap mapped on galGal6 with a bin size of 2.5kb (chr7:15790000-16700000). The positions of the two TADs (T-DOM and C-DOM) and of both sub-TADs are shown on top. The scales on the X axes were adjusted to comparable sizes for ease of comparison, yet the chicken locus is more compacted (see the scale bars, bottom right).

Figure 3. Tissue- and gene-specific interactions of dense H3K27ac regions. A. On top is a ChIP-seq profile of H3K27ac using mouse E12.5 VPs (orange) superimposed over the E12.5 forebrain cells (grey) (mm10, chr2:73800000-75800000). Below is a H3K27ac profile produced from dissected chicken HH35 dorsal skin (green), with HH18 brain cells (grey)(galGal6, chr7:15790000-16700000. The highest interacting regions are depicted as D1, D4 and D9 (see Figure S3B). The positions of conserved CNEs were used to delimit these regions in both species (vertical dashed lines) and the extents of TADs are shown below (thick black lines).

The region in T-DOM containing 30 percent of the total T-DOM acetylation (Most Acetylated Region; MAR) is shown as a red rectangle. In the mouse VP, the MAR is within the D1 DNA segment, whereas the chicken MAR overlaps with the D4 and D9 regions. **B.** H3K27ac ChIP-seq using mouse E9.5 posterior trunk cells (orange) superimposed over forebrain cells (grey)(mm10, chr2:73800000-75800000. Below is shown the H3K27ac profile obtained from the same sample but dissected from a HH18 chicken embryo (green) with brain cells (grey) as a control (galGal6, chr7:15790000-16700000). While the mouse PT Most Acetylated Region (MAR, red rectangle) coincides with the D4 segment, the chicken MAR counterpart also covers D4 and a small part of D9.

Figure 4. Regulation of *Hoxd1*. A. WISH using a *Hoxd1* probe on E12.5 mouse embryos. **B**. X-gal staining on E12.5 mouse embryos carrying randomly integrated BACs. **C**. The three deletion lines used are shown (red) as well as the transgenic BACs (blue). The extent of the D1 to D9 regions are shown on top (black) as well as the positions of both sub-TADs below (thick black lines). *Hoxd1* mRNAs were detected in both the control and the Del(SB2-SB3) lines (**A**, (red arrowheads), yet the signal was absent from Del(Attp-SB2) mutant embryos (**A**, white arrowheads). For transgenics, a strong staining was detected in VPs when BAC^{Mtx2} was used (**B**, red arrowhead), whereas staining was not scored with the other two flanking BAC clones (**B**, white arrowheads). Therefore, a region directly downstream the *HoxD* cluster is necessary and sufficient to activate *Hoxd1* transcription in VPs, referred to as Vibrissa Primordium Regulatory Region (**C**, VPRR). **D.** Enlargement of the VPRR region along with a H3K27ac ChIP-seq using dissected E12.5 VPs (orange) superimposed over E12.5 forebrain (FB) cells (grey)(mm10; chr2:74765140-74913171). Above the profiles are the four VPAEs (open arrowheads) and below are MACS2 narrowPeaks for VPs (orange) and FB (grey). The position of the four sequences containing VPAEs and used as transgenes (tgVPAEs) are shown by numbered black boxes (bottom, Figure S5C). **E.** X-gal staining in VPs on a E12.5 embryo transgenic for VPAE2. **F.** WISH of *Hoxd1* on E9.5 embryos focusing on tail buds. *Hoxd1* was detected in forming somites (FS) in both the control and the Del(SB2-SB3) lines (red arrowheads), but was absent from Del(Attp-SB2) mutant embryos (white arrowhead). **G.** X-gal staining of E9.5 mouse embryos carrying various BAC transgenes. Staining was scored in the entire neural tube and paraxial mesoderm with the BAC^{HoxD} (black arrowhead). In contrast, embryos carrying BAC^{Mtx2} displayed staining in forming somites (red arrowhead), a staining absent when BAC^{T1} was used (white arrowhead). **H.** H3K27ac ChIP-seq over the VPRR region using E9.5 PT cells (orange) superimposed over E12.5 FB cells (grey). Above the profiles are the two SAE sequences (black arrowheads) and below are MACS2 narrowPeaks for PT (orange) and FB (grey) cells. The position of the SAE sequence used as a transgene is shown as a red box (tgSAE). **I.** X-gal staining in forming somites on E9.5 embryo transgenic for the VPAE3 sequence.

Figure 5. Conserved sequences with divergent functions. A. Bar plots showing the proportion of putative *cis*-regulatory elements (pCREs) harboring the H3K27ac mark in one ('specific') *versus* more than one ('pleiotropic') tissues, in either mouse (orange) or chicken (green). For simplification purposes, distal pCREs (>2kb away from the transcription start site) are referred as 'enhancers' and proximal pCREs (<2kb away from

transcription start site) are called 'promoters'. The different proportions are shwon when comparing enhancers and promoters (left), and conserved enhancers to non-conserved enhancers (right). The proportion of specific *versus* pleiotropic elements is shown for conserved and non-conserved enhancers (right). In both mouse and chicken, enhancers are more specific than promoters and amongst enhancers, conserved elements are found acetylated in more tissues than non-conserved elements. **B.** Heatmap obtained using pheatmap R package. On the *X*-axis are shown enhancer-CNEs i.e., conserved sequences, which are non-coding in mouse and in chicken and which overlap with a H3K27ac mark in at least one species. On the *Y*-axis are displayed those tissues used to obtain MACS2 processed peaks. Tissue-specific activity (H3K27ac-positive) is shown in red, whereas abence of H3K27ac is in blue. FB, forebrain; PFL, proximal forelimb; DFL, distal forelimb; PT, posterior trunk; DS, dorsal skin. CNEs cluster exactly by species. **C.** Euler plots representing the proportion of enhancer-CNEs, per tissue, acetylated in mouse only (orange), chicken only (green), or both (pistachio). In **B** and **C,** FB, forebrain; PL, proximal limb; DL, distal limb; PT, posterior trunk; DS, dorsal skin.

LEGENDS TO SUPPLEMENTARY FIGURES

Supplementary Figure S1. Evolution of the *HoxD* locus and expression of *Hoxd* genes in the developing VPs. A. On the left is a schematic illustration of a vertebrate phylogenetic tree showing the emergence of some skin derivatives (glands, scales, feathers, hairs). On the right are the corresponding versions of the *HoxD* locus, with the orthologous *HoxD* gene clusters as shaded orange areas and other orthologous genes shown in green to help delimit the extent of the conserved regulatory landscapes. **B.** Schematic of the transversal section on E13.5 VPs (left), analyzed in ISH with a *Hoxd1* probe (right). The scale bar in the right picture corresponds to 100 µm and the *Hoxd1* stained VPs at peg stage (see below) are indicated by red arrowheads of decreasing intensities. The staining is localized in dermal cells. **C.** WISH on E12.5 bisected mouse embryos. *Hoxd1* mRNAs were detected after those for *Shh* (left) and before *Hoxd4* mRNAs (right). **D.** Schematic representation of vibrissae morphogenesis: epithelial placode (EP, yellow cells); dermal condensation (DC, blue cells). The approximate developmental stage is indicated below, showing the delay between vibrissae and hair morphogenesis. The dorsal skin of the upper back was used as reference for hairs. **E.** WISH on three successive E12 staged embryos (indicated on the left) using four *Hoxd* genes as probes (top). The same three placodes are pointed by red (positive) or white (negative).

Supplementary Figure S2. Feather development and transcription of *Hoxd* genes. A. Schematic of feather development between stage HH20 and HH35 with ectodermal placode (EP, yellow) and dermal condensation cells (DC, blue). **B**. Transversal plan through a HH35 chicken skin of the upper back (red), corresponding to the two sections stained by ISH on the right. The scale bar is 100 µm. *SHH* is used as marker of the epithelial placode (red arrowhead) and *HOXD8* is transcribed in epithelial placodes, matching -or slightly larger than- the *SHH* domain. **C.** WISH on HH35 chicken embryos, dorsal view of the upper back, with arrowheads pointing toward the neck at the shoulder level. *HOXD11* transcripts were not detected in FPs, unlike those of SHH.

Supplementary Figure S3. Comparative 4C-seq analyses at the mouse and chicken *HoxD* loci. A. Synteny plots representing sequences conserved between the mouse and the chicken *HoxD* loci. On the X-axis is the mouse locus (mm10, chr2:73800000-75800000) and on the Y axis is the chicken locus (galGal6, chr7:15790000-16700000). The scale is the same for both axes. TADs and sub-TADs are indicated by black lines and the *HoxD* cluster by a red rectangle. Schematic representations of the locus are on top of each axis. Despite a mouse locus that is in average 2.2 times larger than its chicken counterpart, the order of the conserved sequences is maintained, showing the absence of substantial genomic rearrangement at this locus. **B**. On the Y axis are 4C-seq normalised scores per feature, with E9.5 mouse PT cells scores (orange) superimposed over E12.5 FB cells (grey). Below the 4C tracks, the positions of the *Hoxd1, Hoxd4* and *Hoxd9* viewpoints are represented by triangles, with a black triangle for the viewpoint used in the track. Between each track, the subtraction of the two adjacent viewpoints is represented (black). The preferentially interacting regions D1, D4 and D9 were annotated manually and, to facilitate comparisons between their positions in the mouse and the chicken locus, orthologous CNEs located at the edge of each region D are shown (vertical dashed lines). CTCF ChIP-seq profiles are displayed below the 4C tracks, with red and blue arrowheads below indicating their orientation. Color intensity is proportional to motif score. Data mapped on mm10, chr2:73800000-75800000. **C.** On the Y axis are 4C-seq normalised scores per feature, using HH18 chicken PT cells (green), superimposed with the profile obtained with HH18 brain cells (grey). Data shown for *HOXD1* (top) *HOXD4* (middle) and *HOXD9* (bottom) viewpoints, with a black arrowhead indicating the position of the viewpoint. Reads mapped on galGal6, chr7:15790000-16700000.

Supplementary Figure S4. Comparison of conserved and non-conserved CREs in mouse and chicken embryonic tissues. A. H3K27ac ChIP-seq profiles enlarged over the mouse and chicken VPRR (mm10, chr2:74775737-75222876). Top: Mouse E12.5 dissected VPs (orange) along with the mouse E12.5 forebrain cells track (grey). Below is the profile for the chicken HH35 dorsal skin (green), superimposed to HH18 brain cells (grey) (GalGal6, chr7:16033612-16252401). Processed MACS2 narrowPeaks are in orange/green vertical lines and BLAST (reference) conserved sequences in black vertical lines, below each track. The peaks overlapping CNEs are pointed by black arrows. The peaks corresponding to CNEs in both species and in equivalent tissues are pointed by red arrows. Vertical dashed lines represent the position of corresponding CNEs at the edge of each region **B.** H3K27ac ChIP-seq focusing on VPRR. Top: The profile for mouse E9.5 posterior trunk cells (orange) are superimposed to mouse E12.5 forebrain cells (grey). Bottom: Profiles of chicken HH18 embryonic posterior trunk cells (green) and of HH18 brain cells.

Supplementary Figure S5. *Hoxd1* regulation in mouse vibrissae primordia (VP) and posterior trunk (PT) cells. A. Along the Y-axis are 4C-seq normalised score per feature using E12.5 dissected mouse VPs (orange), superimposed (braun) over E12.5 brain cells as control (grey). Below the 4C tracks, the positions of the *Hoxd1, Hoxd4* and *Hoxd9* viewpoints are represented by triangles with the viewpoint used in the track in black. Between each track is the subtraction of PT to VPs scores (dark braun). The data are mapped on

mm10 (chr2:73800000-75800000). The positions of the D1, D4 and D9 regions are shown on top and the TADs are indicated with thick black lines at the bottom. The vertical dashed lines correspond to orthologous CNEs used for positioning the various regions **B.** VISTA alignments of sequences used as transgenes, to visualize nucleotide conservation. Species are on the left, sequences on top. **C.** X-gal staining of E12.5 embryos carrying transgenes that are not reporting any regulatory activity in VPs (tg^{VPAE1}, tg^{VPAE3}, tg^{VPAE4}) **D.** S-gal staining of a E9.5 mouse embryo transgenic for the tg^{SAE} sequence, showing no staining in forming somites.

Supplementary Figure S6. Lineage-specific H3K27ac-positive elements active in chicken FPAEs. A. ChIP H3K27ac profile using either dissected chicken HH35 dorsal skin (green), or HH18 brain cells (grey) over the MAR region defined in Figure 3. **B.** VISTA alignments of sequences specifically enriched for the H327ac epigenetic mark in the embryonic chicken skin (HH35), but not in the brain and posterior trunk samples, and located within the FPAR. Species are on the left, sequences on top. No significant homology of these elements was found within the mouse *HoxD* genomic region.

LEGENDS TO SUPPLEMENTARY TABLES

Supplementary Table S1. List of mouse lines

Supplementary Table S2. Primers used to clone candidate enhancer sequences

Supplementary Table S3. Probes for *in situ* hybridization

BIBLIOGRAPHY

Acemel, R.D., Tena, J.J., Irastorza-Azcarate, I., Marlétaz, F., Gómez-Marín, C., de la Calle-Mustienes, E., Bertrand, S., Diaz, S.G., Aldea, D., Aury, J.-M., Mangenot, S., Holland, P.W.H., Devos, D.P., Maeso, I., Escrivá, H., Gómez-Skarmeta, J.L., 2016. A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. Nat. Genet. 48, 336–341. https://doi.org/10.1038/ng.3497

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., Grüning, B., Guerler, A., Hillman-Jackson, J., Von Kuster, G., Rasche, E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A., Goecks, J., 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 44, W3–W10. https://doi.org/10.1093/nar/gkw343

Allais-Bonnet, A., Hintermann, A., Deloche, M.-C., Cornette, R., Bardou, P., Naval-Sanchez, M., Pinton, A., Haruda, A., Grohs, C., Zakany, J., Bigi, D., Medugorac, I., Putelat, O., Greyvenstein, O., Hadfield, T., Jemaa, S.B., Bunevski, G., Menzi, F., Hirter, N., Paris, J.M., Hedges, J., Palhiere, I., Rupp, R., Lenstra, J.A., Gidney, L., Lesur, J., Schafberg, R., Stache, M., Wandhammer, M.-D., Arbogast, R.-M., Guintard, C., Blin, A., Boukadiri, A., Rivière, J., Esquerré, D., Donnadieu, C., Danchin-Burge, C., Reich, C.M., Riley, D.G., Marle-Koster, E. van, Cockett, N., Hayes, B.J., Drögemüller, C., Kijas, J., Pailhoux, E., Tosser-Klopp, G., Duboule, D., Capitan, A., 2021. Analysis of Polycerate Mutants Reveals the Evolutionary Co-option of *HOXD1* for Horn Patterning in Bovidae. Mol. Biol. Evol. 38, 2260–2272. https://doi.org/10.1093/molbev/msab021

Amândio, A.R., Lopez-Delisle, L., Bolt, C.C., Mascrez, B., Duboule, D., 2020. A complex regulatory landscape involved in the development of mammalian external genitals. eLife 9, e52962. https://doi.org/10.7554/eLife.52962

Ambrosino, L., Vassalli, Q.A., D'Agostino, Y., Esposito, R., Cetrangolo, V., Caputi, L., Amoroso, A., Aniello, F., D'Aniello, S., Chatzigeorgiou, M., Chiusano, M.L., Locascio, A., 2019. Functional conserved non-coding elements among tunicates and chordates. Dev. Biol., Current Directions in Tunicate Development 448, 101–110. https://doi.org/10.1016/j.ydbio.2018.12.012

Andrey, G., Montavon, T., Mascrez, B., Gonzalez, F., Noordermeer, D., Leleu, M., Trono, D., Spitz, F., Duboule, D., 2013. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. Science 340, 1234167. https://doi.org/10.1126/science.1234167

Beccari, L., Yakushiji-Kaminatsui, N., Woltering, J.M., Necsulea, A., Lonfat, N., Rodriguez-Carballo, E., Mascrez, B., Yamamoto, S., Kuroiwa, A., Duboule, D., 2016. A role for HOX13 proteins in the regulatory switch between TADs at the HoxD locus. Genes Dev. 30, 1172–1186. https://doi.org/10.1101/gad.281055.116

Belton, J.-M., McCord, R.P., Gibcus, J., Naumova, N., Zhan, Y., Dekker, J., 2012. Hi-C: A comprehensive technique to capture the conformation of genomes. Methods San Diego Calif 58. https://doi.org/10.1016/j.ymeth.2012.05.001

Berthelot, C., Villar, D., Horvath, J.E., Odom, D.T., Flicek, P., 2018. Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. Nat. Ecol. Evol. 2, 152–163. https://doi.org/10.1038/s41559-017-0377-2

Biggs, L.C., Mikkola, M.L., 2014. Early inductive events in ectodermal appendage morphogenesis.

Bolt, C.C., Duboule, D., 2020. The regulatory landscapes of developmental genes. Development 147, dev171736. https://doi.org/10.1242/dev.171736

Bolt, C.C., Lopez-Delisle, L., Mascrez, B., Duboule, D., 2021. MESOMELIC DYSPLASIAS ASSOCIATED WITH THE HOXD LOCUS ARE CAUSED BY REGULATORY REALLOCATIONS. bioRxiv 2021.02.01.429171. https://doi.org/10.1101/2021.02.01.429171

Carroll, S.B., 2008. Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. Cell 134, 25–36. https://doi.org/10.1016/j.cell.2008.06.030

Chiang, C., Swan, R.Z., Grachtchouk, M., Bolinger, M., Litingtung, Y., Robertson, E.K., Cooper, M.K., Gaffield, W., Westphal, H., Beachy, P.A., Dlugosz, A.A., 1999. Essential role for Sonic

hedgehog during hair follicle morphogenesis. Dev. Biol. 205, 1–9. https://doi.org/10.1006/dbio.1998.9103
Chuong, C.M., Oliver, G., Ting, S.A., Jegalian, B.G., Chen, H.M., De Robertis, E.M., 1990. Gradients of homeoproteins in developing feather buds. Dev. Camb. Engl. 110, 1021–1030.
Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., Boyer, L.A., Young, R.A., Jaenisch, R., 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci USA 107, 21931–21936. https://doi.org/10.1073/pnas.1016071107
Dale, K.J., Pourquié, O., 2000. A clock-work somite. BioEssays 22, 72–83. https://doi.org/10.1002/(SICI)1521-1878(200001)22:1<72::AID-BIES12>3.0.CO;2-S
Darbellay, F., Bochaton, C., Lopez-Delisle, L., Mascrez, B., Tschopp, P., Delpretti, S., Zakany, J., Duboule, D., 2019. The constrained architecture of mammalian Hox gene clusters. Proc. Natl. Acad. Sci. U. S. A. 116, 13424–13433. https://doi.org/10.1073/pnas.1904602116
Darbellay, F., Duboule, D., 2016. Chapter Sixteen - Topological Domains, Metagenes, and the Emergence of Pleiotropic Regulations at Hox Loci, in: Wassarman, P.M. (Ed.), Current Topics in Developmental Biology, Essays on Developmental Biology, Part A. Academic Press, pp. 299–314.
David, F.P.A., Delafontaine, J., Carat, S., Ross, F.J., Lefebvre, G., Jarosz, Y., Sinclair, L., Noordermeer, D., Rougemont, J., Leleu, M., 2014. HTSstation: A Web Application and Open-Access Libraries for High-Throughput Sequencing Data Analysis. PLoS ONE 9, e85879. https://doi.org/10.1371/journal.pone.0085879
Dekker, J., Heard, E., 2015. Structural and functional diversity of Topologically Associating Domains. FEBS Lett. 589, 2877–2884. https://doi.org/10.1016/j.febslet.2015.08.044
Delpretti, S., Montavon, T., Leleu, M., Joye, E., Tzika, A., Milinkovitch, M., Duboule, D., 2013. Multiple enhancers regulate Hoxd genes and the Hotdog LncRNA during cecum budding. Cell Rep. 5, 137–50. https://doi.org/10.1016/j.celrep.2013.09.002
Dermitzakis, E.T., Clark, A.G., 2002. Evolution of Transcription Factor Binding Sites in Mammalian Gene Regulatory Regions: Conservation and Turnover. Mol. Biol. Evol. 19, 1114–1121. https://doi.org/10.1093/oxfordjournals.molbev.a004169
Dhouailly, D., 2009. A new scenario for the evolutionary origin of hair, feather, and avian scales. J. Anat. 214, 587–606. https://doi.org/10.1111/j.1469-7580.2008.01041.x
Dhouailly, D., Godefroit, P., Martin, T., Nonchev, S., Caraguel, F., Oftedal, O., 2019. Getting to the root of scales, feather and hair: As deep as odontodes? Exp. Dermatol. 28, 503–508. https://doi.org/10.1111/exd.13391
Di-Poï, N., Milinkovitch, M.C., 2016. The anatomical placode in reptile scale morphogenesis indicates shared ancestry among skin appendages in amniotes. Sci. Adv. 2, e1600708. https://doi.org/10.1126/sciadv.1600708
Dixon, J.R., Gorkin, D.U., Ren, B., 2016. Chromatin Domains: The Unit of Chromosome Organization. Mol. Cell 62, 668–680. https://doi.org/10.1016/j.molcel.2016.05.018
Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., Ren, B., 2012. Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. Nature 485, 376–380. https://doi.org/10.1038/nature11082
Edwards, S.L., Beesley, J., French, J.D., Dunning, A.M., 2013. Beyond GWASs: Illuminating the Dark Road from Association to Function. Am. J. Hum. Genet. 93, 779–797. https://doi.org/10.1016/j.ajhg.2013.10.012
Eichenlaub, M.P., Ettwiller, L., 2011. De Novo Genesis of Enhancers in Vertebrates. PLoS Biol. 9. https://doi.org/10.1371/journal.pbio.1001188
Eres, I.E., Luo, K., Hsiao, C.J., Blake, L.E., Gilad, Y., 2019. Reorganization of 3D genome structure may contribute to gene regulatory evolution in primates. PLOS Genet. 15, e1008278. https://doi.org/10.1371/journal.pgen.1008278
Fernandez-Guerrero, M., Yakushiji-Kaminatsui, N., Lopez-Delisle, L., Zdral, S., Darbellay, F., Perez-Gomez, R., Bolt, C.C., Sanchez-Martin, M.A., Duboule, D., Ros, M.A., 2020. Mammalian-

specific ectodermal enhancers control the expression of Hoxc genes in developing nails and hair follicles. Proc. Natl. Acad. Sci. U. S. A. 117, 30509–30519. https://doi.org/10.1073/pnas.2011078117
Gaunt, S.J., 1988. Mouse homeobox gene transcripts occupy different but overlapping domains in embryonic germ layers and organs: a comparison of Hox-3.1 and Hox-1.5. Development 103, 135–44.
Godwin, A.R., Capecchi, M.R., 1998. Hoxc13 mutant mice lack external hair. Genes Dev. 12, 11–20.
Guerreiro, I., Gitto, S., Novoa, A., Codourey, J., Nguyen Huynh, T.H., Gonzalez, F., Milinkovitch, M.C., Mallo, M., Duboule, D., 2016. Reorganisation of Hoxd regulatory landscapes during the evolution of a snake-like body plan. eLife 5, e16087. https://doi.org/10.7554/eLife.16087
Hamburger, V., Hamilton, H.L., 1951. A series of normal stages in the development of the chick embryo. J. Morphol. 88, 49–92. https://doi.org/10.1002/jmor.1050880104
Hansen, A.S., Pustova, I., Cattoglio, C., Tjian, R., Darzacq, X., 2017. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. eLife 6, e25776. https://doi.org/10.7554/eLife.25776
Harmston, N., Ing-Simmons, E., Tan, G., Perry, M., Merkenschlager, M., Lenhard, B., 2017. Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. Nat. Commun. 8. https://doi.org/10.1038/s41467-017-00524-5
Holland, L.Z., Albalat, R., Azumi, K., Benito-Gutiérrez, È., Blow, M.J., Bronner-Fraser, M., Brunet, F., Butts, T., Candiani, S., Dishaw, L.J., Ferrier, D.E.K., Garcia-Fernàndez, J., Gibson-Brown, J.J., Gissi, C., Godzik, A., Hallböök, F., Hirose, D., Hosomichi, K., Ikuta, T., Inoko, H., Kasahara, M., Kasamatsu, J., Kawashima, T., Kimura, A., Kobayashi, M., Kozmik, Z., Kubokawa, K., Laudet, V., Litman, G.W., McHardy, A.C., Meulemans, D., Nonaka, M., Olinski, R.P., Pancer, Z., Pennacchio, L.A., Pestarino, M., Rast, J.P., Rigoutsos, I., Robinson-Rechavi, M., Roch, G., Saiga, H., Sasakura, Y., Satake, M., Satou, Y., Schubert, M., Sherwood, N., Shiina, T., Takatori, N., Tello, J., Vopalensky, P., Wada, S., Xu, A., Ye, Y., Yoshida, K., Yoshizaki, F., Yu, J.-K., Zhang, Q., Zmasek, C.M., Jong, P.J. de, Osoegawa, K., Putnam, N.H., Rokhsar, D.S., Satoh, N., Holland, P.W.H., 2008. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. Genome Res. 18, 1100–1111. https://doi.org/10.1101/gr.073676.107
Holland, P.W.H., Garcia-Fernàndez, J., 1996. HoxGenes and Chordate Evolution. Dev. Biol. 173, 382–395. https://doi.org/10.1006/dbio.1996.0034
Irimia, M., Tena, J.J., Alexis, M.S., Fernandez-Miñan, A., Maeso, I., Bogdanovic, O., de la Calle-Mustienes, E., Roy, S.W., Gómez-Skarmeta, J.L., Fraser, H.B., 2012. Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. Genome Res. 22, 2356–2367. https://doi.org/10.1101/gr.139725.112
Kanzler, B., Prin, F., Thelu, J., Dhouailly, D., 1997. CHOXC-8 and CHOXD-13 expression in embryonic chick skin and cutaneous appendage specification. Dev. Dyn. Off. Publ. Am. Assoc. Anat. 210, 274–287. https://doi.org/10.1002/(SICI)1097-0177(199711)210:3<274::AID-AJA8>3.0.CO;2-D
Kanzler, B., Viallet, J.P., Le Mouellic, H., Boncinelli, E., Duboule, D., Dhouailly, D., 1994. Differential expression of two different homeobox gene families during mouse tegument morphogenesis. Int. J. Dev. Biol. 38, 633–640.
Kragesteen, B.K., Spielmann, M., Paliou, C., Heinrich, V., Schöpflin, R., Esposito, A., Annunziatella, C., Bianco, S., Chiariello, A.M., Jerković, I., Harabula, I., Guckelberger, P., Pechstein, M., Wittler, L., Chan, W.-L., Franke, M., Lupiáñez, D.G., Kraft, K., Timmermann, B., Vingron, M., Visel, A., Nicodemi, M., Mundlos, S., Andrey, G., 2018. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. Nat. Genet. 50, 1463–1473. https://doi.org/10.1038/s41588-018-0221-x
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–9. https://doi.org/10.1038/nmeth.1923

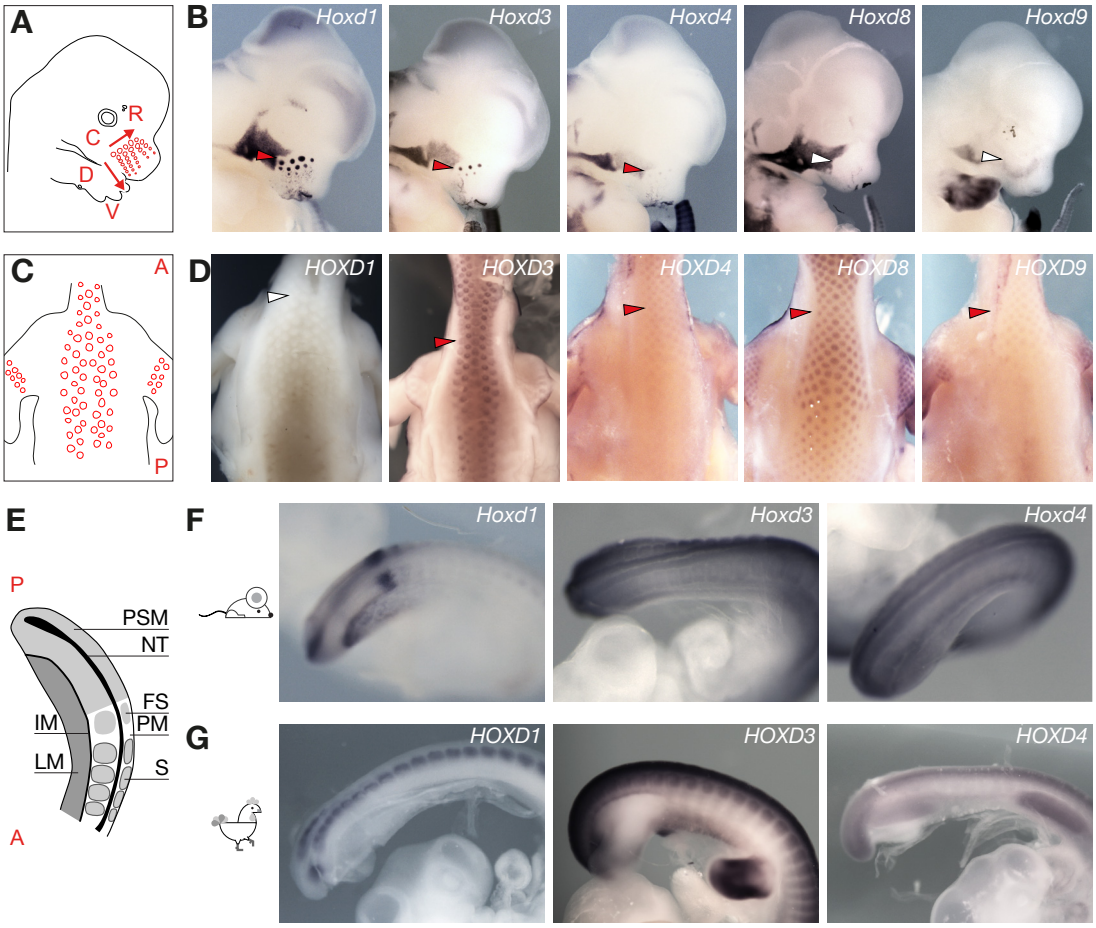
Laverré, A., Tannier, E., Necsulea, A., 2021. Long-range promoter-enhancer contacts are conserved during evolution and contribute to gene expression robustness. bioRxiv 2021.02.26.432473. https://doi.org/10.1101/2021.02.26.432473
Liao, Y., Zhang, X., Chakraborty, M., Emerson, J.J., 2021. Topologically associating domains and their role in the evolution of genome structure and function in Drosophila. Genome Res. 31, 397–410. https://doi.org/10.1101/gr.266130.120
Long, H.K., Prescott, S.L., Wysocka, J., 2016. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell 167, 1170–1187. https://doi.org/10.1016/j.cell.2016.09.018
Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S.A., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., Spielmann, M., Visel, A., Mundlos, S., 2015. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. Cell 161, 1012–1025. https://doi.org/10.1016/j.cell.2015.04.004
Marinic, M., Aktas, T., Ruf, S., Spitz, F., 2013. An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. Dev. Cell 24, 530–42. https://doi.org/10.1016/j.devcel.2013.01.025
Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17. https://doi.org/10.14806/ej.17.1.200
Michon, F., Charveron, M., Dhouailly, D., 2007. Dermal condensation formation in the chick embryo: Requirement for integrin engagement and subsequent stabilization by a possible Notch/integrin interaction. Dev. Dyn. 236, 755–768. https://doi.org/10.1002/dvdy.21080
Mikkola, M.L., 2007. Genetic basis of skin appendage development. Semin. Cell Dev. Biol., Mechanisms and biological significance of RNA localization AND Development of ectodermal appendages 18, 225–236. https://doi.org/10.1016/j.semcdb.2007.01.007
Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F., Duboule, D., 2011. A Regulatory Archipelago Controls Hox Genes Transcription in Digits. Cell 147, 1132–1145. https://doi.org/10.1016/j.cell.2011.10.023
Nanni, L., Ceri, S., Logie, C., 2020. Spatial patterns of CTCF sites define the anatomy of TADs and their boundaries. Genome Biol. 21, 197. https://doi.org/10.1186/s13059-020-02108-x
Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., De Laat, W., Duboule, D., 2011. The dynamic architecture of Hox gene clusters. Science 334, 222–5. https://doi.org/10.1126/science.1207194
Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., Berkum, N.L. van, Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J., Heard, E., 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485, 381. https://doi.org/10.1038/nature11049
Packer, A.I., Jane-wit, D., McLean, L., Panteleyev, A.A., Christiano, A.M., Wolgemuth, D.J., 2000. Hoxa4 expression in developing mouse hair follicles and skin. Mech. Dev. 99, 153–157. https://doi.org/10.1016/S0925-4773(00)00471-8
Pantalacci, S., Chaumot, A., Benoît, G., Sadier, A., Delsuc, F., Douzery, E.J.P., Laudet, V., 2008. Conserved features and evolutionary shifts of the EDA signaling pathway involved in vertebrate skin appendage development. Mol. Biol. Evol. 25, 912–928. https://doi.org/10.1093/molbev/msn038
Pugacheva, E.M., Kubo, N., Loukinov, D., Tajmul, M., Kang, S., Kovalchuk, A.L., Strunnikov, A.V., Zentner, G.E., Ren, B., Lobanenkov, V.V., 2020. CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. Proc. Natl. Acad. Sci. U. S. A. 117, 2020–2031. https://doi.org/10.1073/pnas.1911708117
Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., Wysocka, J., 2011. A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470, 279–283. https://doi.org/10.1038/nature09692

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., Aiden, E.L., 2014. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell 159, 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021
Rebeiz, M., Patel, N.H., Hinman, V.F., 2015. Unraveling the Tangled Skein: The Evolution of Transcriptional Regulatory Networks in Development. Annu. Rev. Genomics Hum. Genet. 16, 103–131. https://doi.org/10.1146/annurev-genom-091212-153423
Reid, A.I., Gaunt, S.J., 2002. Colinearity and non-colinearity in the expression of Hox genes in developing chick skin. Int. J. Dev. Biol. 46, 209–15. https://doi.org/10.1387/ijdb.011495
Reynolds, A.J., Lawrence, C.M., Jahoda, C.A., 1995. Hair follicle and fiber reconstruction. J. Invest. Dermatol. 104, 39S-40S. https://doi.org/10.1038/jid.1995.57
Rodríguez-Carballo, E., Lopez-Delisle, L., Willemin, A., Beccari, L., Gitto, S., Mascrez, B., Duboule, D., 2020. Chromatin topology and the timing of enhancer function at the HoxD locus. Proc. Natl. Acad. Sci. U. S. A. 117, 31231–31241. https://doi.org/10.1073/pnas.2015083117
Rodriguez-Carballo, E., Lopez-Delisle, L., Zhan, Y., Fabre, P.J., Beccari, L., El-Idrissi, I., Huynh, T.H.N., Ozadam, H., Dekker, J., Duboule, D., 2017. The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. Genes Dev. 31, 2264–2281. https://doi.org/10.1101/gad.307769.117
Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., Li, C.H., Guo, Y.E., Day, D.S., Schuijers, J., Vasile, E., Malik, S., Hnisz, D., Lee, T.I., Cisse, I.I., Roeder, R.G., Sharp, P.A., Chakraborty, A.K., Young, R.A., 2018. Coactivator condensation at super-enhancers links phase separation and gene control. Science 361. https://doi.org/10.1126/science.aar3958
Sandelin, A., Bailey, P., Bruce, S., Engström, P.G., Klos, J.M., Wasserman, W.W., Ericson, J., Lenhard, B., 2004. Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. BMC Genomics 5, 99. https://doi.org/10.1186/1471-2164-5-99
Sanges, R., Hadzhiev, Y., Gueroult-Bellone, M., Roure, A., Ferg, M., Meola, N., Amore, G., Basu, S., Brown, E.R., De Simone, M., Petrera, F., Licastro, D., Strähle, U., Banfi, S., Lemaire, P., Birney, E., Müller, F., Stupka, E., 2013. Highly conserved elements discovered in vertebrates are present in non-syntenic loci of tunicates, act as enhancers and can be transcribed during development. Nucleic Acids Res. 41, 3600–3618. https://doi.org/10.1093/nar/gkt030
Sanges, R., Kalmar, E., Claudiani, P., D'Amato, M., Muller, F., Stupka, E., 2006. Shuffling of cis-regulatory elements is a pervasive feature of the vertebrate lineage. Genome Biol. 7, R56. https://doi.org/10.1186/gb-2006-7-7-r56
Sawyer, R.H., Knapp, L.W., 2003. Avian skin development and the evolutionary origin of feathers. J. Exp. Zoolog. B Mol. Dev. Evol. 298, 57–72. https://doi.org/10.1002/jez.b.26
Schep, R., Necsulea, A., Rodríguez-Carballo, E., Guerreiro, I., Andrey, G., Nguyen Huynh, T.H., Marcet, V., Zákány, J., Duboule, D., Beccari, L., 2016. Control of Hoxd gene transcription in the mammary bud by hijacking a preexisting regulatory landscape. Proc. Natl. Acad. Sci. U. S. A. 113, E7720–E7729. https://doi.org/10.1073/pnas.1617141113
Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S., Talianidis, I., Flicek, P., Odom, D.T., 2010. Five-Vertebrate ChIP-seq Reveals the Evolutionary Dynamics of Transcription Factor Binding. Science 328, 1036–1040. https://doi.org/10.1126/science.1186176
Shao, Z., Zhang, Y., Yuan, G.-C., Orkin, S.H., Waxman, D.J., 2012. MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. Genome Biol. 13, R16. https://doi.org/10.1186/gb-2012-13-3-r16
Snetkova, V., Ypsilanti, A.R., Akiyama, J.A., Mannion, B.J., Plajzer-Frick, I., Novak, C.S., Harrington, A.N., Pham, Q.T., Kato, M., Zhu, Y., Godoy, J., Meky, E., Hunter, R.D., Shi, M., Kvon, E.Z., Afzal, V., Tran, S., Rubenstein, J.L.R., Visel, A., Pennacchio, L.A., Dickel, D.E., 2021. Ultraconserved enhancer function does not require perfect sequence conservation. Nat. Genet. 53,

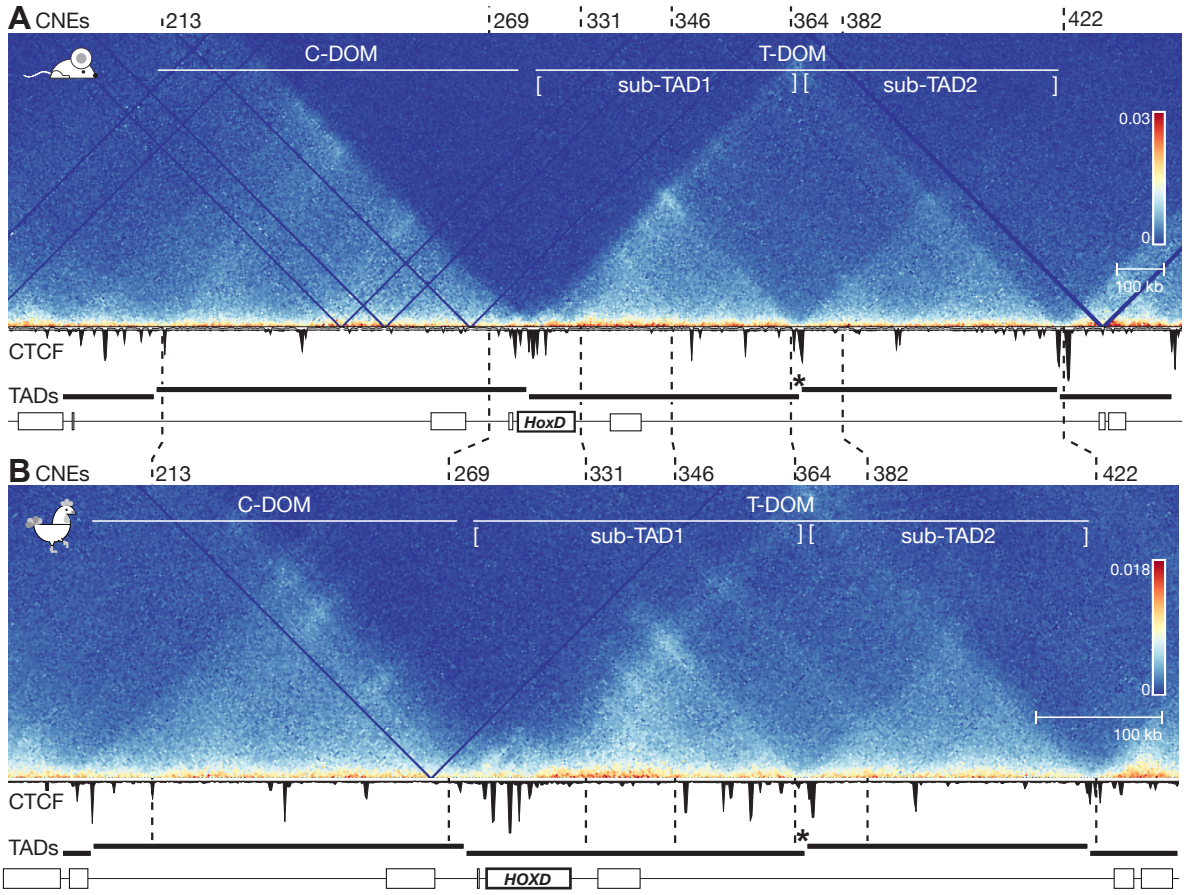
521–528. https://doi.org/10.1038/s41588-021-00812-3
Sockanathan, S., Sockanathan, S., 2015. In situ hybridization of chick and mouse embryonic tissue. Protoc. Exch. https://doi.org/10.1038/protex.2015.040
Spitz, F., Furlong, E.E.M., 2012. Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. 13, 613–626. https://doi.org/10.1038/nrg3207
Spitz, F., Gonzalez, F., Peichel, C., Vogt, T.F., Duboule, D., Zakany, J., 2001. Large scale transgenic and cluster deletion analysis of the HoxD complex separate an ancestral regulatory module from evolutionary innovations. Genes Dev 15, 2209–14. https://doi.org/10.1101/gad.205701
Torosin, N.S., Anand, A., Golla, T.R., Cao, W., Ellison, C.E., 2020. 3D genome evolution and reorganization in the Drosophila melanogaster species group. PLOS Genet. 16, e1009229. https://doi.org/10.1371/journal.pgen.1009229
Ushiki, A., Zhang, Y., Xiong, C., Zhao, J., Georgakopoulos-Soares, I., Kane, L., Jamieson, K., Bamshad, M.J., Nickerson, D.A., University of Washington Center for Mendelian Genomics, Shen, Y., Lettice, L.A., Silveira-Lucas, E.L., Petit, F., Ahituv, N., 2021. Deletion of CTCF sites in the SHH locus alters enhancer-promoter interactions and leads to acheiropodia. Nat. Commun. 12, 2282. https://doi.org/10.1038/s41467-021-22470-z
Vierstra, J., Rynes, E., Sandstrom, R., Zhang, M., Canfield, T., Hansen, R.S., Stehling-Sun, S., Sabo, P.J., Byron, R., Humbert, R., Thurman, R.E., Johnson, A.K., Vong, S., Lee, K., Bates, D., Neri, F., Diegel, M., Giste, E., Haugen, E., Dunn, D., Wilken, M.S., Josefowicz, S., Samstein, R., Chang, K.-H., Eichler, E.E., De Bruijn, M., Reh, T.A., Skoultchi, A., Rudensky, A., Orkin, S.H., Papayannopoulou, T., Treuting, P.M., Selleri, L., Kaul, R., Groudine, M., Bender, M.A., Stamatoyannopoulos, J.A., 2014. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. Science 346, 1007–1012. https://doi.org/10.1126/science.1246426
Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., Turner, J.M.A., Bertelsen, M.F., Murchison, E.P., Flicek, P., Odom, D.T., 2015. Enhancer Evolution across 20 Mammalian Species. Cell 160, 554–566. https://doi.org/10.1016/j.cell.2015.01.006
Widelitz, R.B., Chuong, C.M., 1999. Early events in skin appendage formation: induction of epithelial placodes and condensation of dermal mesenchyme. J. Investig. Dermatol. Symp. Proc. Soc. Investig. Dermatol. Inc Eur. Soc. Dermatol. Res. 4, 302–306. https://doi.org/10.1038/sj.jidsp.5640234
Woltering, J.M., Noordermeer, D., Leleu, M., Duboule, D., 2014. Conservation and Divergence of Regulatory Strategies at Hox Loci and the Origin of Tetrapod Digits. PLoS Biol 12, e1001773. https://doi.org/10.1371/journal.pbio.1001773
Wong, E.S., Zheng, D., Tan, S.Z., Bower, N.I., Garside, V., Vanwalleghem, G., Gaiti, F., Scott, E., Hogan, B.M., Kikuchi, K., McGlinn, E., Francois, M., Degnan, B.M., 2020. Deep conservation of the enhancer regulatory code in animals. Science 370. https://doi.org/10.1126/science.aax8137
Wu, P., Hou, L., Plikus, M., Hughes, M., Scehnet, J., Suksaweang, S., Widelitz, R., Jiang, T.-X., Chuong, C.-M., 2004. Evo-Devo of amniote integuments and appendages. Int. J. Dev. Biol. 48, 249–270. https://doi.org/10.1387/ijdb.041825pw
Yakushiji-Kaminatsui, N., Lopez-Delisle, L., Bolt, C.C., Andrey, G., Beccari, L., Duboule, D., 2018. Similarities and differences in the regulation of HoxD genes during chick and mouse limb development. PLoS Biol. 16, e3000004. https://doi.org/10.1371/journal.pbio.3000004
Yu, Z., Jiang, K., Xu, Z., Huang, H., Qian, N., Lu, Z., Chen, D., Di, R., Yuan, T., Du, Z., Xie, W., Lu, X., Li, H., Chai, R., Yang, Y., Zhu, B., Kunieda, T., Wang, F., Chen, T., 2018. Hoxc-Dependent Mesenchymal Niche Heterogeneity Drives Regional Hair Follicle Regeneration. Cell Stem Cell 23, 487-500.e6. https://doi.org/10.1016/j.stem.2018.07.016
Zákány, J., Kmita, M., Alarcon, P., de la Pompa, J.-L., Duboule, D., 2001. Localized and Transient Transcription of Hox Genes Suggests a Link between Patterning and the Segmentation Clock. Cell 106, 207–217. https://doi.org/10.1016/S0092-8674(01)00436-6

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S., 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol 9, R137. https://doi.org/10.1186/gb-2008-9-9-r137
Ziebarth, J.D., Bhattacharya, A., Cui, Y., 2012. CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. Nucleic Acids Res. 41, D188–D194. https://doi.org/10.1093/nar/gks1165

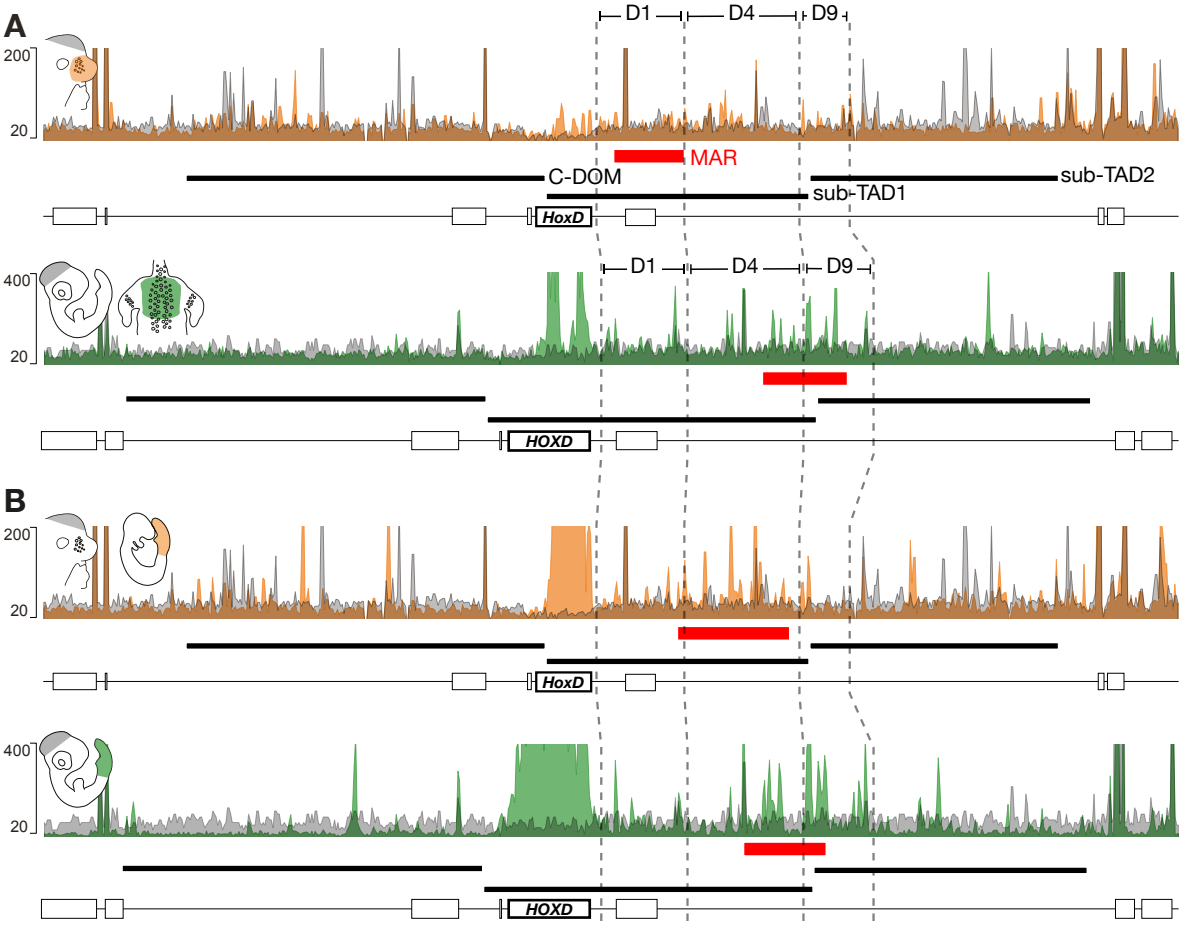
Hintermann et al.
Figure 1

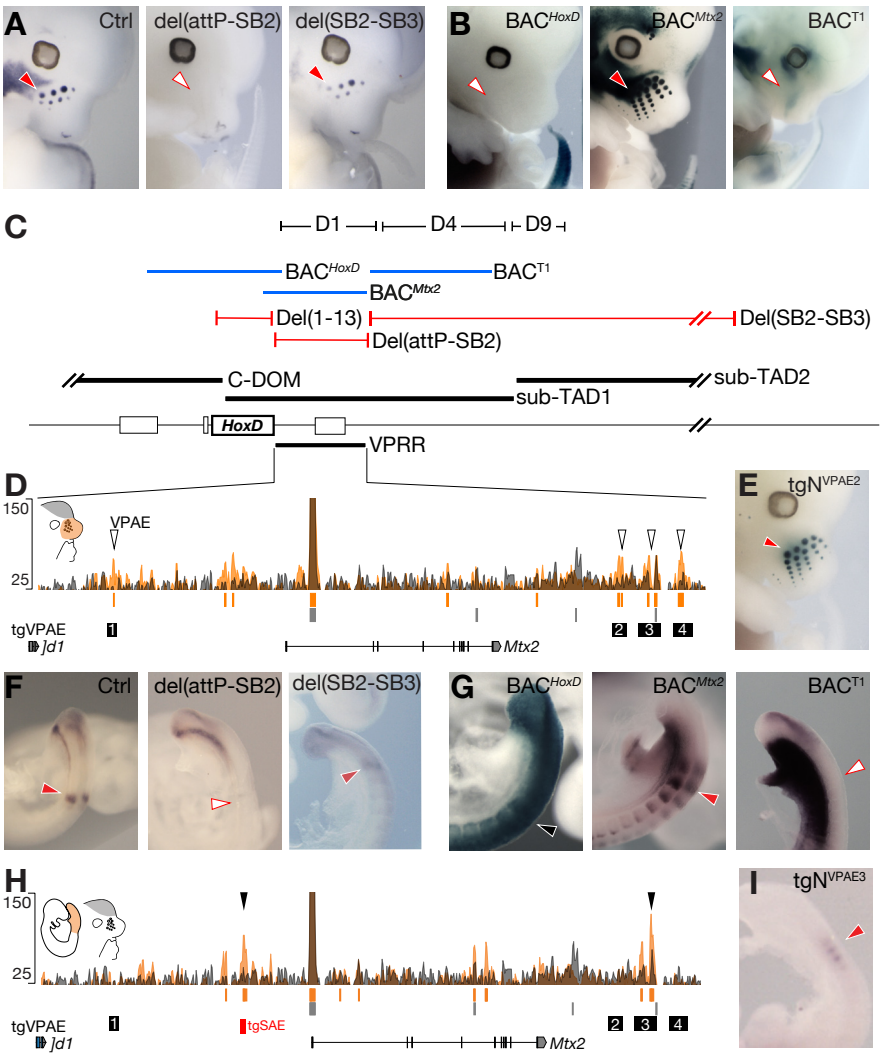


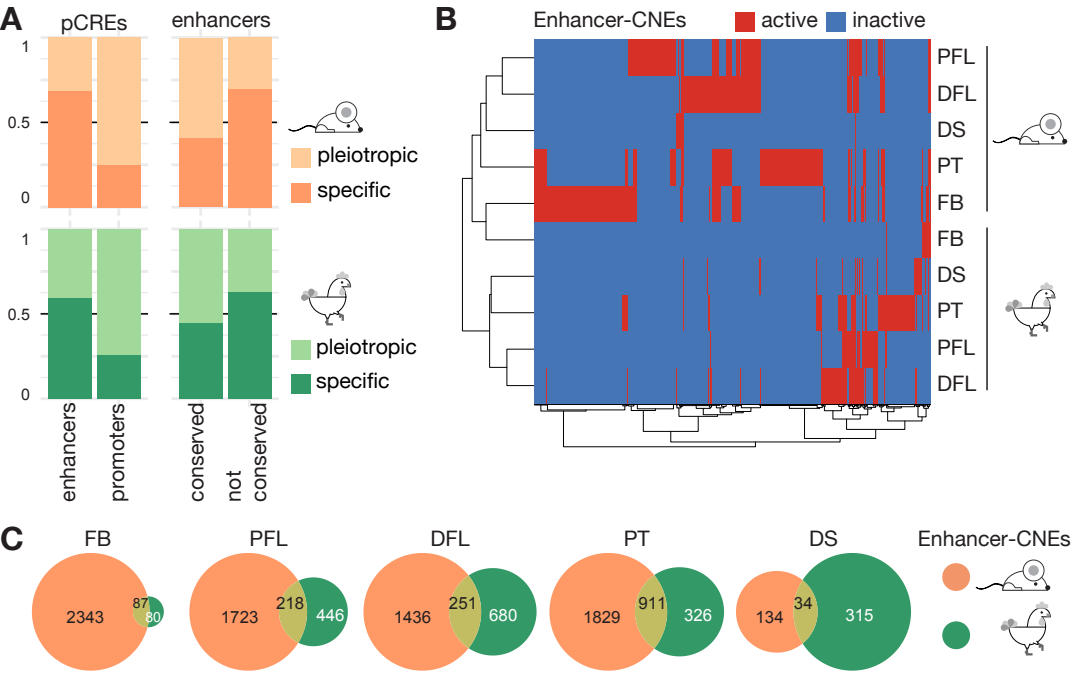
Hintermann et al.
Figure 2



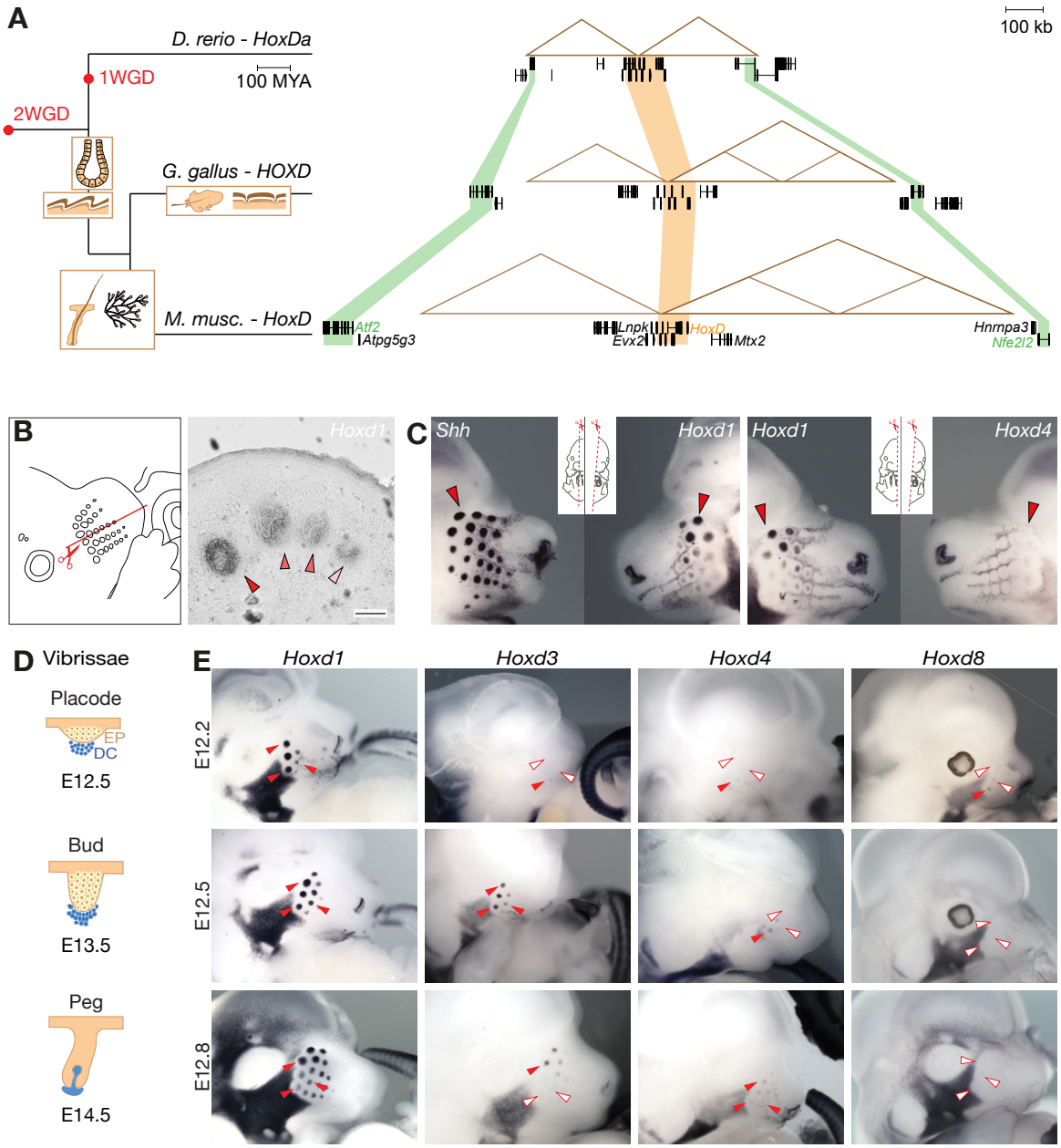
Hintermann et al.
Figure 3

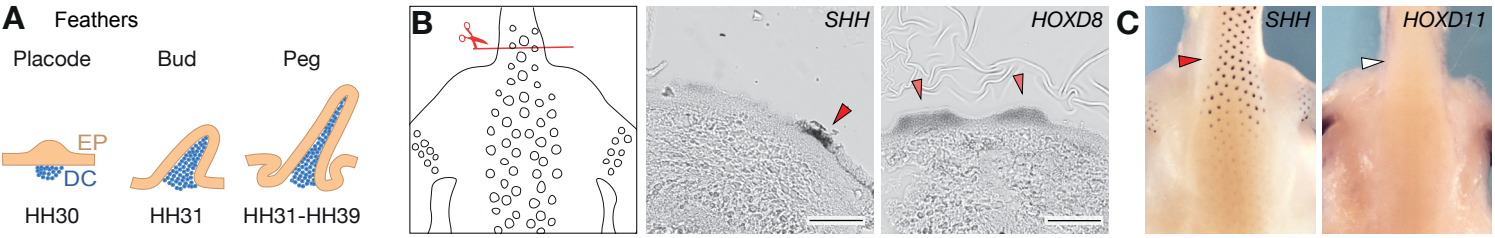


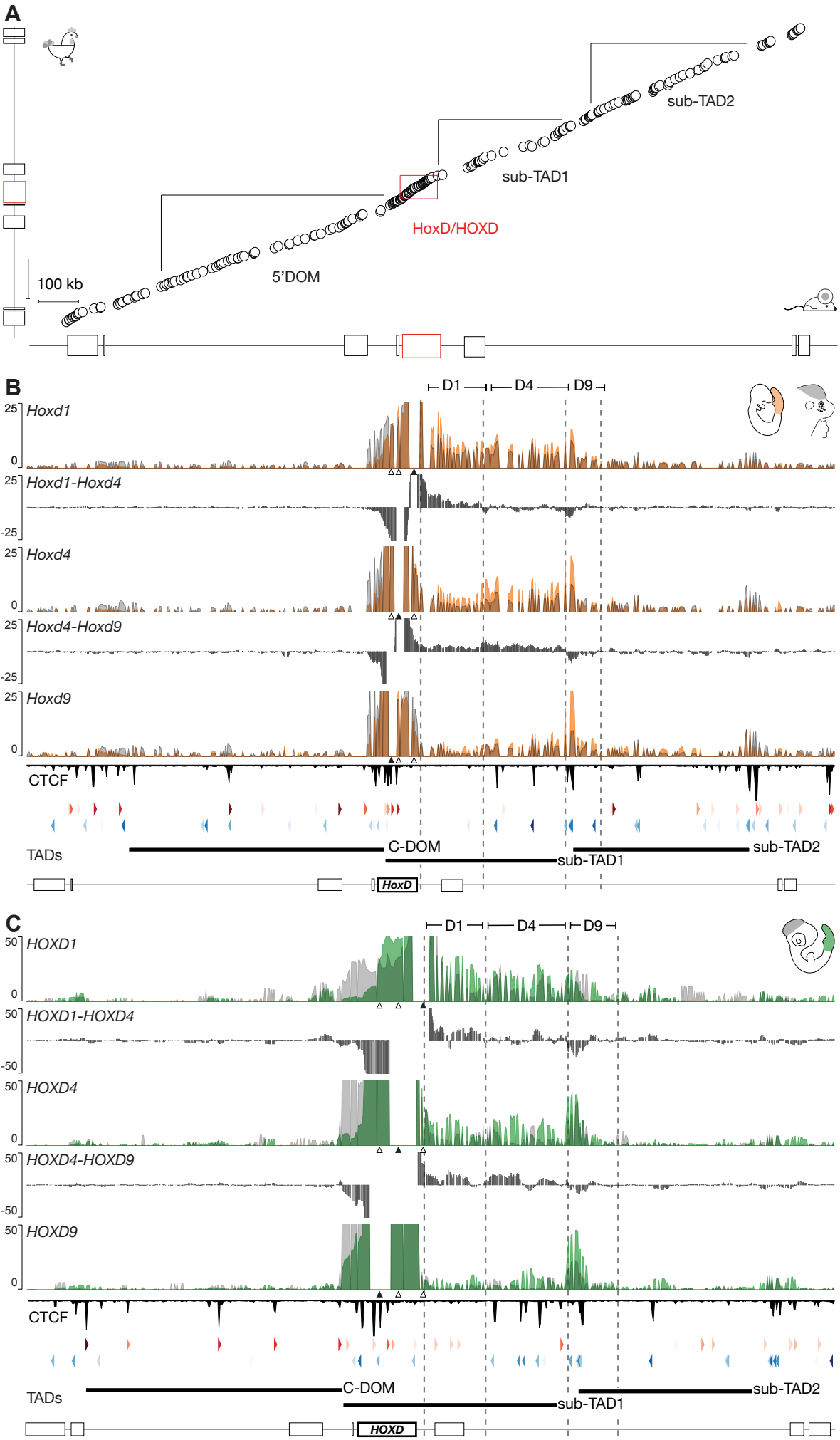

A
Ctrl
del(attP-SB2)
del(SB2-SB3)
B
BAC^HoxD
BAC^Mtx2
BAC^T1
C
D1
D4
D9
BAC^HoxD
BAC^T1
BAC^Mtx2
Del(1-13)
Del(SB2-SB3)
Del(attP-SB2)
C-DOM
sub-TAD2
sub-TAD1
HoxD
VPRR
D
150
25
VPAE
tgVPAE
Jd1
1
2
3
4
Mtx2
E
tgN^VPAE2
F
Ctrl
del(attP-SB2)
del(SB2-SB3)
G
BAC^HoxD
BAC^Mtx2
BAC^T1
H
150
25
tgVPAE
Jd1
1
tgSAE
2
3
4
Mtx2
I
tgN^VPAE3

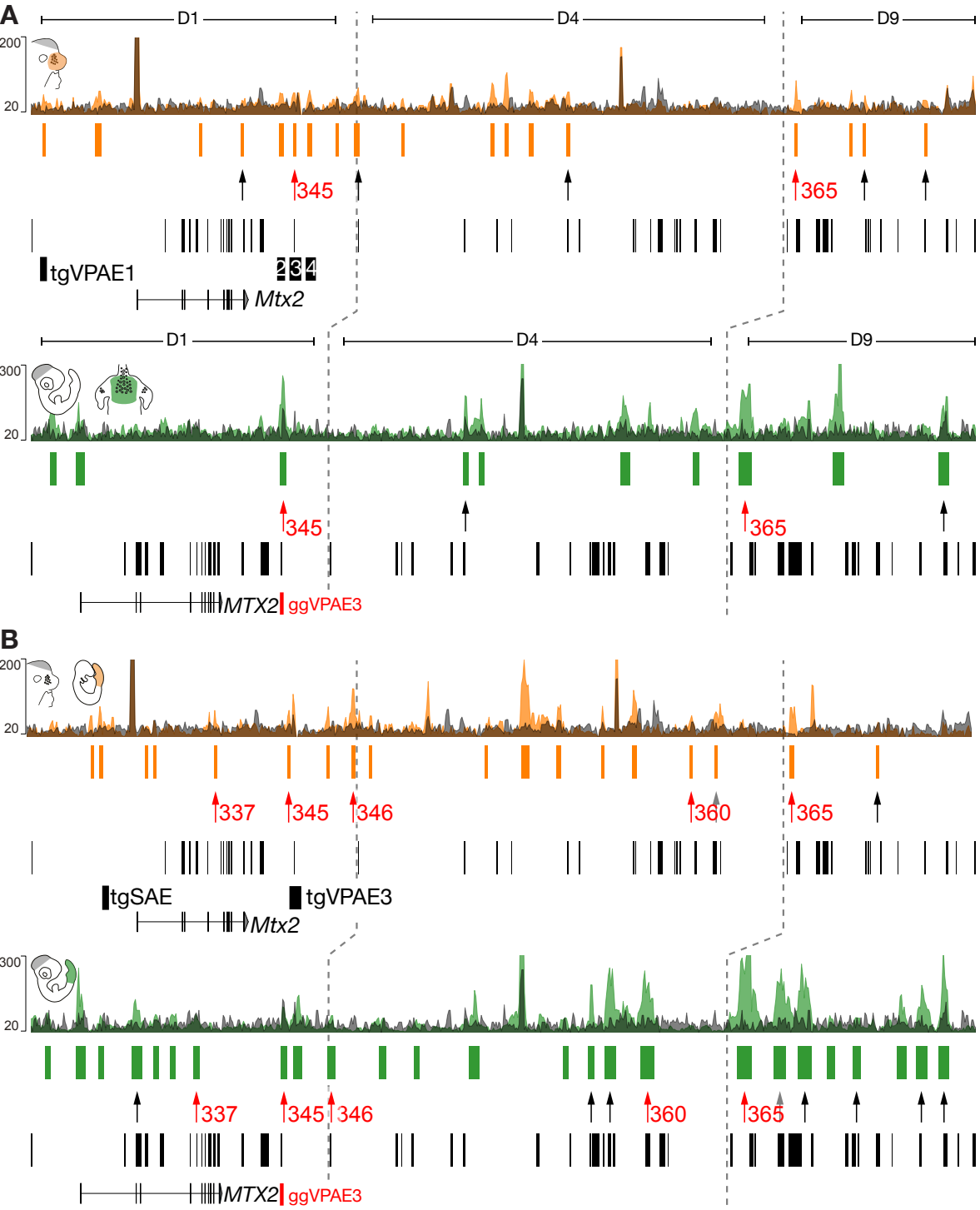

A
pCREs
enhancers
1
0.5
0
pleiotropic
specific
enhancers
promoters
conserved
not conserved
B
Enhancer-CNEs
active
inactive
PFL
DFL
DS
PT
FB
FB
DS
PT
PFL
DFL

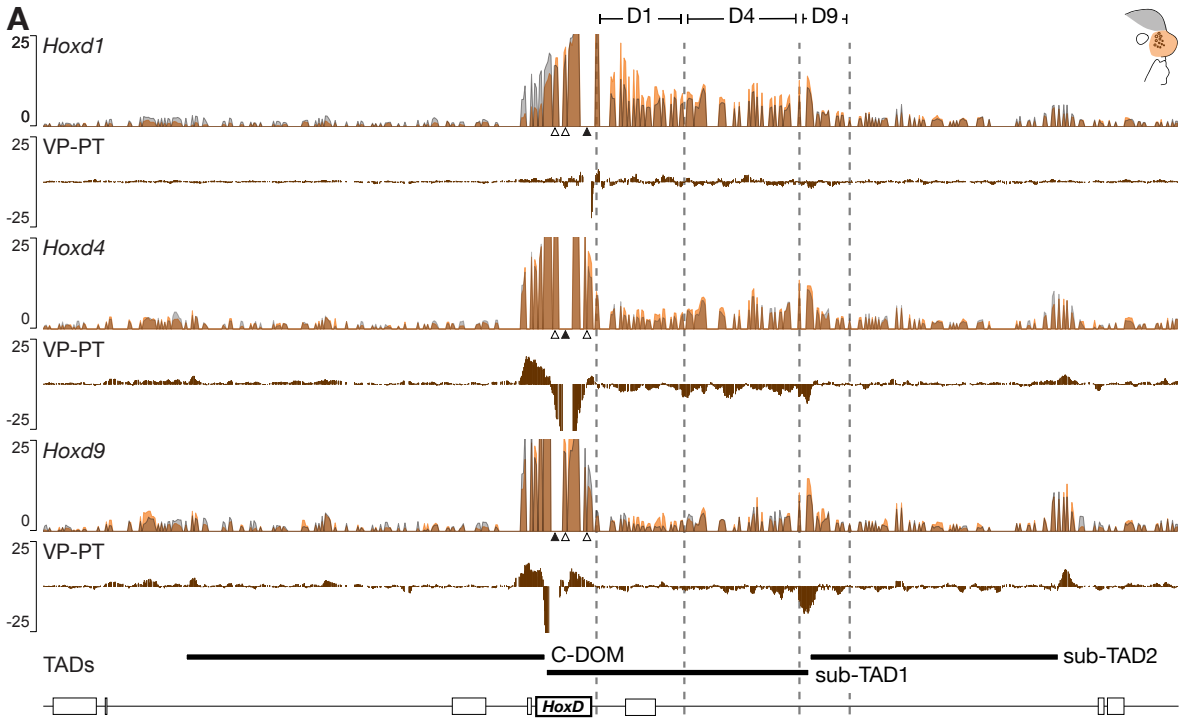
C
FB
PFL
DFL
PT
DS
Enhancer-CNEs
2343
87
80
1723
218
446
1436
251
680
1829
911
326
134
34
315

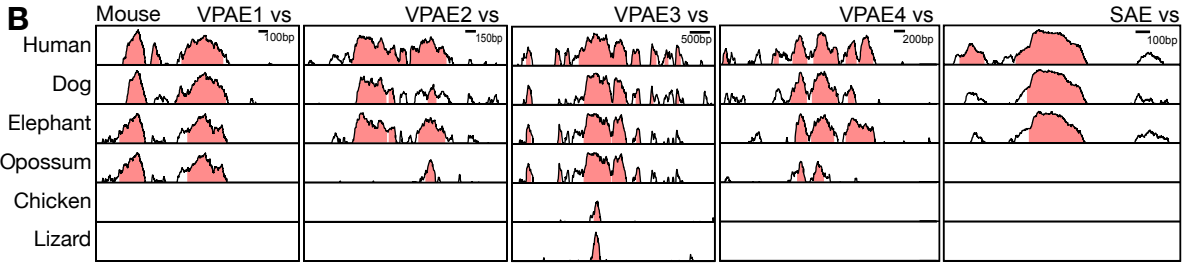

A
D. rerio - HoxDa
1WGD
100 MYA
2WGD
G. gallus - HOXD
M. musc. - HoxD
100 kb
Atf2
Atpg5g3
Lnpk
Evx2
HoxD
Mtx2
Hnrnpa3
Nfe2l2
B
Hoxd1
C
Shh
Hoxd1
Hoxd1
Hoxd4
D Vibrissae
Placode
EP
DC
E12.5
Bud
E13.5
Peg
E14.5
E
Hoxd1
Hoxd3
Hoxd4
Hoxd8
E12.2
E12.5
E12.8

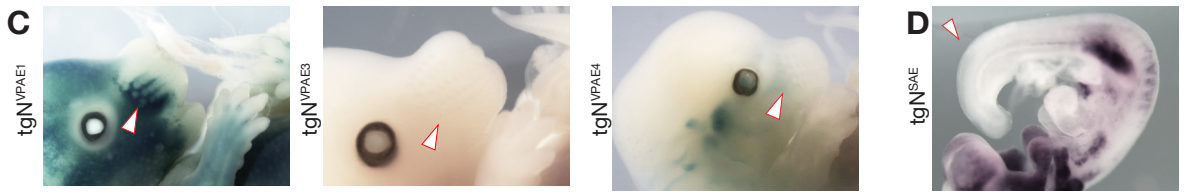

A
Feathers
Placode
Bud
Peg
EP
DC
HH30
HH31
HH31-HH39
B
SHH
HOXD8
C
SHH
HOXD11

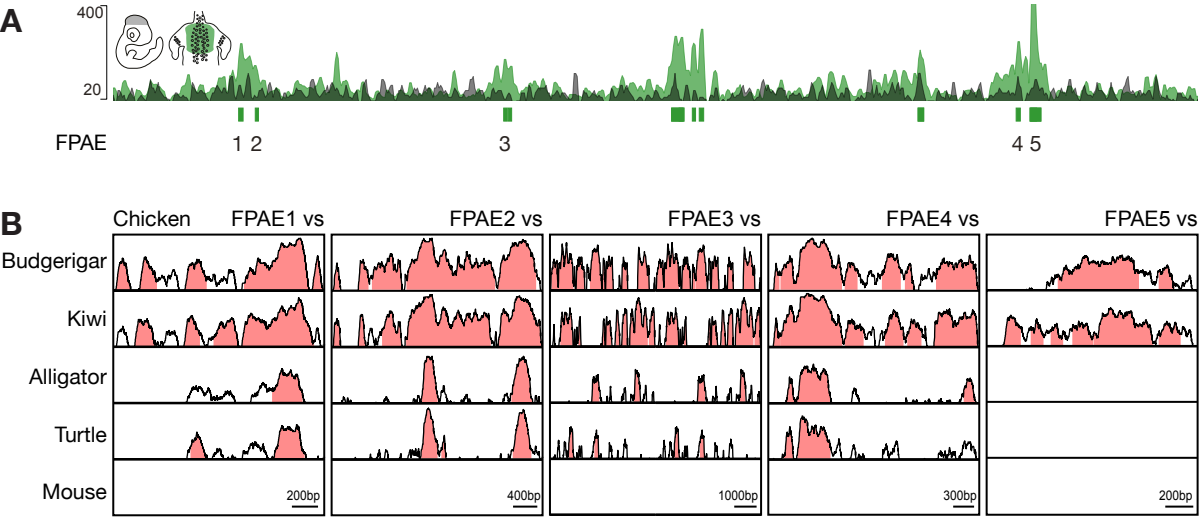

A
sub-TAD2
sub-TAD1
HoxD/HOXD
5'DOM
100 kb
B
D1
D4
D9
Hoxd1
Hoxd1-Hoxd4
Hoxd4
Hoxd4-Hoxd9
Hoxd9
CTCF
TADs
C-DOM
sub-TAD2
sub-TAD1
HoxD
C
HOXD1
HOXD1-HOXD4
HOXD4
HOXD4-HOXD9
HOXD9
CTCF
TADs
C-DOM
sub-TAD2
sub-TAD1
HOXD


A
D1
D4
D9
200
20
345
365
tgVPAE1
2 3 4
Mtx2
D1
D4
D9
300
20
345
365
MTX2
ggVPAE3
B
200
20
337
345
346
360
365
tgSAE
tgVPAE3
Mtx2
300
20
337
345
346
360
365
MTX2
ggVPAE3


A
D1
D4
D9
Hoxd1
25
0
VP-PT
25
-25
Hoxd4
25
0
VP-PT
25
-25
Hoxd9
25
0
VP-PT
25
-25
TADs
C-DOM
sub-TAD2
sub-TAD1
HoxD


B
Mouse
VPAE1 vs
VPAE2 vs
VPAE3 vs
VPAE4 vs
SAE vs
100bp
150bp
500bp
200bp
100bp
Human
Dog
Elephant
Opossum
Chicken
Lizard


C
tgN(VPAE1)
tgN(VPAE3)
tgN(VPAE4)
D
tgN(SAE)


A
400
20
FPAE
1 2
3
4 5
B
Chicken
FPAE1 vs
FPAE2 vs
FPAE3 vs
FPAE4 vs
FPAE5 vs
Budgerigar
Kiwi
Alligator
Turtle
Mouse
200bp
400bp
1000bp
300bp
200bp

CONCLUDING REMARKS

Altogether, our results support the conceptual framework proposing that TADs, long-range enhancers and compact gene clusters are the three components of a virtuous evolutionary cycle, which outputs increased pleiotropy of developmental genes in vertebrates and associated phenotypical diversity. In this view, chromatin architecture and gene regulation co-evolve and are intimately related.

We show that a pre-existing domain can evolve lineage specific regulatory modalities in lineage specific structures, or alternatively, achieve transcriptional robustness in ancestrally derived structures despite a drastic loss of sequence conservation. This genomic set up, where *cis*-regulatory elements evolve rapidly within an overall conserved topological domain, may provide a balance between transcriptional stability and regulatory evolvability.

If TADs are dispensable for most enhancer-promoter communications, we propose that they are essential for establishment of new regulation relationships during evolution.

ANNEX I

Analysis of Polycerate Mutants Reveals the Evolutionary Co-option of *HOXD1* for Horn Patterning in Bovidae

Aurélie Allais-Bonnet,[†,1,2,3] Aurélie Hintermann,[†,4] Marie-Christine Deloche,[1,2,3] Raphaël Cornette,[5] Philippe Bardou,[6,7] Marina Naval-Sanchez,[8] Alain Pinton,[6] Ashleigh Haruda,[9] Cécile Grohs,[10] Jozsef Zakany,[4] Daniele Bigi,[11] Ivica Medugorac,[12] Olivier Putelat,[13,14] Ockert Greyvenstein,[15] Tracy Hadfield,[16] Slim Ben Jemaa,[17] Gjoko Bunevski,[18] Fiona Menzi,[19] Nathalie Hirter,[19] Julia M. Paris,[19] John Hedges,[20] Isabelle Palhiere,[6] Rachel Rupp,[6] Johannes A. Lenstra,[21] Louisa Gidney,[22] Joséphine Lesur,[23] Renate Schafberg,[9] Michael Stache,[9] Marie-Dominique Wandhammer,[24] Rose-Marie Arbogast,[25] Claude Guintard,[26,27] Amandine Blin,[28] Abdelhak Boukadiri,[10] Julie Rivière,[10,29] Diane Esquerré,[30] Cécile Donnadieu,[30] Coralie Danchin-Burge,[31] Coralie M. Reich,[32] David G. Riley,[15] Este van Marle-Koster,[33] Noelle Cockett,[16] Benjamin J. Hayes,[34] Cord Drögemüller,[19] James Kijas,[8] Eric Pailhoux,[2,3] Gwenola Tosser-Klopp,[6] Denis Duboule,[*,4,35,36] and Aurélien Capitan[*,1,10]

[1]ALLICE, Paris, France

[2]Université Paris-Saclay, UVSQ, INRAE, BREED, Jouy-en-Josas, France

[3]Ecole Nationale Vétérinaire d'Alfort, BREED, Maisons-Alfort, France

[4]Department of Genetics and Evolution, University of Geneva, Geneva, Switzerland

[5]Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, Paris, France

[6]GenPhySE, Université de Toulouse, INRAE, ENVT, Castanet-Tolosan, France

[7]INRAE, Sigenae, Castanet-Tolosan, France

[8]CSIRO Agriculture & Food, St. Lucia, QLD, Australia

[9]Central Natural Science Collections, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

[10]Université Paris-Saclay, INRAE, AgroParisTech, GABI, Jouy-en-Josas, France

[11]Dipartimento di Scienza e Tecnologie Agro-Alimentari, Alma Mater Studiorum University of Bologna, Bologna, Italy

[12]Population Genomics Group, Department of Veterinary Sciences, Ludwig-Maximilians-University Munich, Munich, Germany

[13]Archéologie Alsace, Sélestat, France

[14]UMR 7044, ARCHIMEDE, MISHA, Strasbourg, France

[15]Department of Animal Science, Texas A&M University, College Station, TX, USA

[16]Department of Animal, Dairy, and Veterinary Sciences, Utah State University, Logan, UT, USA

[17]Laboratoire des Productions Animales et Fourragères, Institut National de la Recherche Agronomique de Tunisie, Université de Carthage, Ariana, Tunisia

[18]Livestock Department, Faculty of Agricultural Sciences and Food Institute of Animal Biotechnology, University Ss. Cyril and Methodius, Skopje, North Macedonia

[19]Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland

[20]Manx Loaghtan Sheep Breeders' Group, Bassingbourn, Cambridgeshire, United Kingdom

[21]Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands

[22]Rent a Peasant, Tow Law, Bishop Auckland, Durham County, United Kingdom

[23]Unité Archéozoologie, Archéobotanique, Sociétés Pratiques et Environnements (AASPE), CNRS, Muséum National d'Histoire Naturelle, Paris, France

[24]Musée Zoologique de Strasbourg, Strasbourg, France

[25]CNRS UMR 7044, ARCHIMEDE, MISHA, Strasbourg, France

[26]Unité d'Anatomie Comparée, Ecole Nationale Vétérinaire de l'Agroalimentaire et de l'Alimentation, Nantes Atlantique—ONIRIS, Nantes, France

[27]Groupe d'Études Remodelage Osseux et bioMatériaux (GEROM), Université d'Angers, Unité INSERM 922, LHEA/IRIS-IBS, CHU d'Angers, Angers, France

[28]Muséum National d'Histoire Naturelle, CNRS, UMS 2700 2AD, Paris, France

[29]INRAE, Micalis Institute, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

Open Access

Article

[30]INRAE, US, 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, France
[31]Institut de l'Elevage, Paris, France
[32]Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, VIC, Australia
[33]Department of Science, University of Pretoria, Hatfield, South Africa
[34]Queensland Alliance for Agriculture and Food Innovation (QAAFI), Centre for Animal Science, University of Queensland, St. Lucia, QLD, Australia
[35]Swiss Cancer Research Institute, EPFL, Lausanne, Switzerland
[36]Collège de France, Paris, France
[†]Shared first authors.
*Shared last authors and **corresponding authors**: E-mails: aurelien.capitan@inrae.fr; denis.duboule@epfl.ch.
Associate editor: Patricia Wittkopp

Abstract

In the course of evolution, pecorans (i.e., higher ruminants) developed a remarkable diversity of osseous cranial appendages, collectively referred to as "headgear," which likely share the same origin and genetic basis. However, the nature and function of the genetic determinants underlying their number and position remain elusive. Jacob and other rare populations of sheep and goats are characterized by polyceraty, the presence of more than two horns. Here, we characterize distinct *POLYCERATE* alleles in each species, both associated with defective *HOXD1* function. We show that haploinsufficiency at this locus results in the splitting of horn bud primordia, likely following the abnormal extension of an initial morphogenetic field. These results highlight the key role played by this gene in headgear patterning and illustrate the evolutionary co-option of a gene involved in the early development of bilateria to properly fix the position and number of these distinctive organs of Bovidae.

***Key words:* Hox genes, co-option, regulatory mutation, goat and sheep genomics.**

Introduction

In pecorans, successive environmental and behavioral adaptations favored the emergence and sometimes the secondary loss of a variety of headgear, as exemplified by bovid horns, cervid antlers, giraffid ossicones, or antilocaprid pronghorns (Davis et al. 2011; Wang et al. 2019). As different as they are, these iconic organs share both a common cellular origin and a minimal structural organization: they derive from neural crest stem cells and consist of paired structures, located on the frontal bones and composed of a bony core covered by integument (Davis et al. 2011; Wang et al. 2019) (fig. 1 and supplementary fig. 1, Supplementary Material online). Although the development and evolution of headgear is a long-standing question, the underlying molecular and cellular mechanisms have been difficult to study, mostly because the patterning and differentiation of headgear progenitor cells occur early during embryogenesis (Lincoln 1973; Allais-Bonnet et al. 2013) and involve hundreds of genes (Wang et al. 2019).

In this context, natural mutations affecting headgear number, shape, or position, such as the polycerate (multihorned) phenotype occurring in small ruminants (fig. 1*a* and *b*, OMIA entries 000806-9940 and 000806-9925; https://omia.org/home/, last accessed November 4, 2020), offer a valuable alternative (Capitan et al. 2012). Polyceraty was already observed ca 6000 BCE, in the oldest ovine remains from Çatalhöyük, Turkey (Epstein 1971; Putelat 2005) and this dominant trait currently segregates in several sheep breeds around the world. Even though the corresponding locus was mapped in seven distinct populations to the same region of chromosome 2 (Chr2), it has not yet been identified (Greyvenstein et al. 2016; He et al. 2016; Kijas et al. 2016; Ren et al. 2016). In contrast, polycerate goats are observed only sporadically in the Alps. They are not present in archaeological remains and have not been subject to any genetic studies thus far. The oldest record of this condition in goat dates back from 1786, when a four-horned billy-goat was transferred from the city of Bulle in Switzerland to the model farm of French Queen Marie-Antoinette in Versailles (Heitzmann 2006).

In this study, we set up to determine the genetic bases of these conditions in sheep and goats. We show that polyceraty in Bovidae is due either to a 4-bp deletion affecting the splicing of the *HOXD1* gene in sheep, or to the deletion of a large regulatory region controlling the same gene in goats. These results thus illustrate the evolutionary co-option of this gene normally involved in early development to help determine the position and number of horns. They also show that comparable phenotypes observed in distinct species and selected and maintained for a long time are caused by the misregulation of the same gene.

Results and Discussion

Characterization of *POLYCERATE* Mutations in Sheep and Goats

To identify the genetic determinants of polyceraty, we reanalyzed the Illumina OvineHD Beadchip genotyping data (600 k SNPs) of 111 case and 87 control sheeps generated by two

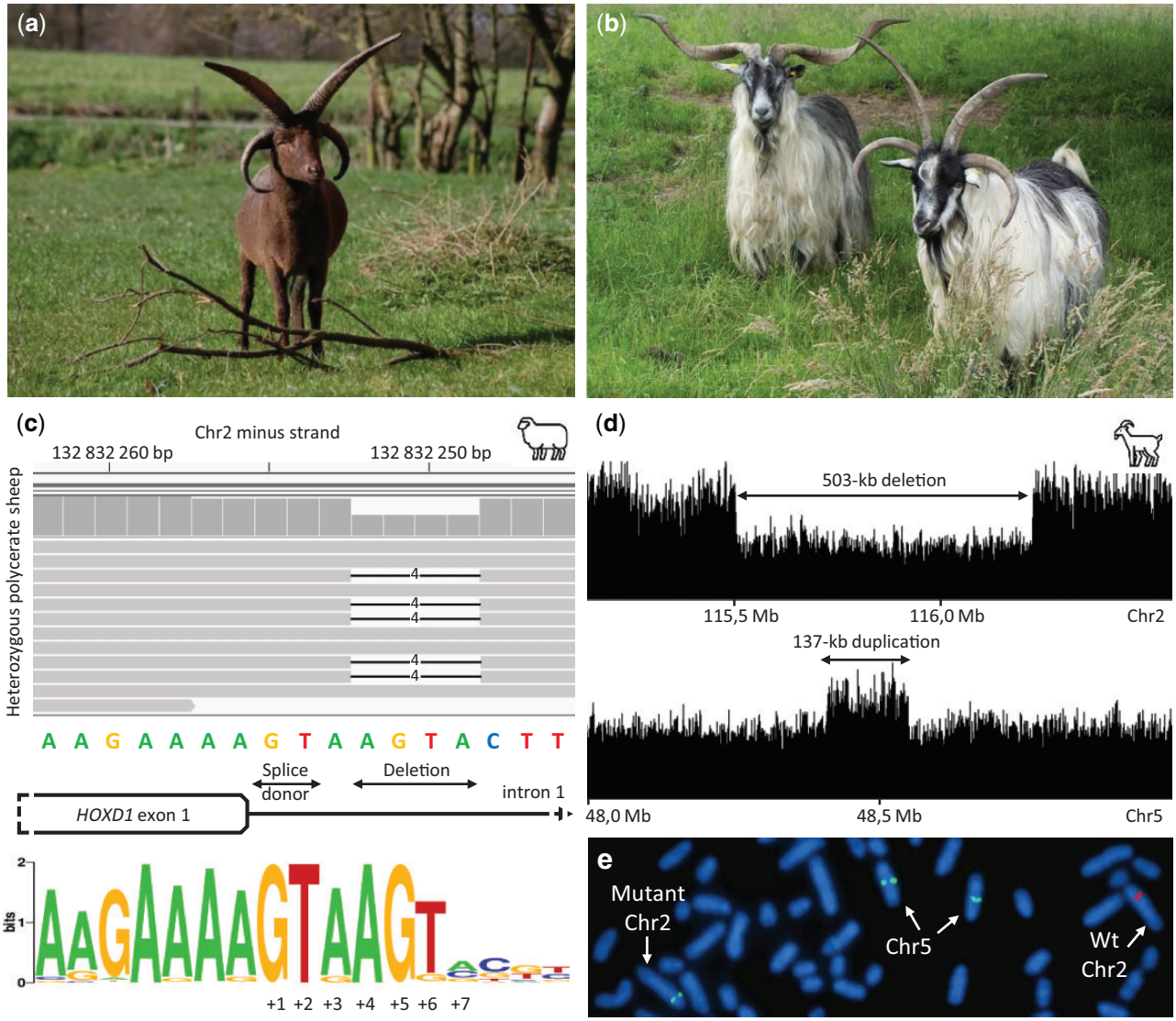


FIG. 1. Polyceraty in sheep and goats and candidate genetic variants. (*a*) Polycerate Manx Loaghtan ram. (*b*) Wild-type and polycerate male goats from a local German population. These individuals represent the most common phenotype. Polycerate animals with asymetric horns and partial fusion of lateral horns are also regularly observed. (*c*) A 4-bp deletion causing polyceraty in sheep. Integrative Genome Viewer (IGV) screenshot with the localization of the variant with respect to HOXD1. Below is a graphical representation of nucleotide conservation at the exon 1-intron junction across 103 sarcopterigian and tetrapod species. (*d*) Plot of read coverage in a heterozygous polycerate goat animal carrying a deletion of 503-kb downstream the *HOXD* gene cluster on Chr2 and a duplication of 137 kb on Chr5. (*e*) FISH-mapping in a heterozygous polycerate goat with BAC clones corresponding to the region deleted in Chr2 (labeled in red) and to the segment of Chr5 inserted at the deletion site (labeled in green). Magnification: ×1,000. Sheep and goat icons were made by "Monkik" from www.thenounproject.com (last accessed November 4, 2020).

previous studies (Greyvenstein et al. 2016; Kijas et al. 2016) (supplementary tables 1 and 2, Supplementary Material online). Assuming autosomal dominant inheritance and genetic homogeneity in the three breeds investigated, we fine-mapped the ovine *POLYCERATE* locus between positions 132,717,593 and 133,151,166 bp on Chr2 (Oar_v4.0 assembly; supplementary fig. 2, Supplementary Material online). By comparing whole-genome sequences of 11 polycerate specimens and 1,179 controls representing the world-wide sheep diversity, we identified a single candidate variant in this interval: a four-nucleotide deletion located at position +4 to +7 bp after exon 1 of the *HOXD1* gene (g.132,832,249_132,832,252del; fig. 1*c*), that is, encompassing three nucleotides (+4, +5, +6) of the consensus splice donor site (Zhang 1998). Genotyping of this variant in 236 animals from eight populations containing polycerate specimens showed a perfect genotype to phenotype association (supplementary tables 3 and 4, Supplementary Material online). Moreover, cross-species alignments revealed that the +4 and +5 nucleotides are conserved among 103 sarcopterigian and tetrapod species, indicating the occurrence of a genuine and consensual splice donor site and hence suggesting a detrimental effect of the micro-deletion in the splicing of *HOXD1* precursor RNAs (fig. 1*c* and supplementary table 5, Supplementary Material online).

We next mapped the caprine *POLYCERATE* mutation to a 542-kb large region orthologous to that of the ovine locus (Chr2:115,143,037–115,685,115 bp on ARS1 assembly; Bickhart et al. 2017; supplementary fig. 2, Supplementary Material online), by using a panel of 35 polycerate and 51 two-horned goats obtained from eight European populations and genotyped with the Illumina GoatSNP50 BeadChip (Tosser-Klopp et al. 2014) (supplementary table 6, Supplementary Material online). Within this interval, we identified 36 private heterozygous variants in one heterozygous polycerate goat versus 1,160 control individuals (supplementary table 7, Supplementary Material online). Genotyping of five case–control pairs from distinct breeds reduced the list of candidates to 15 short variants, affecting genomic regions not conserved among 103 eutherian mammals, as well as a rare type of structural variation located 57 kb downstream of the *HOXD1* 3′-UTR (supplementary tables 7 and 8, Supplementary Material online). The latter involved the translocation of 137 kb from Chr5 to Chr2 by means of a circular intermediate (Durkin et al. 2012) and the deletion of 503 kb from the insertion site (g.115,652,290_116,155,699delins137 kb; fig. 1*d* and *e*), as confirmed by PCR amplification and Sanger sequencing of the regions containing the breakpoints (supplementary fig. 3, Supplementary Material online). Consequently, the mutant chromosome lacked the *MTX2* gene and carried an exogenic copy of both *RASSF3* and the first ten exons of *GNS*. Genotyping of this variant in 77 case and 355 control goats originating from 24 distinct populations revealed a 100 percent association between polyceraty and heterozygosity for the large insertion–deletion (indel, supplementary table 9 and supplementary fig. 4, Supplementary Material online). Homozygous mutants were not detected in our panel, whereas at least 14 polycerate animals were born from polycerate pairs of parents (binomial $P = 3.4 \times 10^{-3}$; supplementary note 1, Supplementary Material online). Because the knockdown of *Mtx2* in zebrafish is embryonic lethal at gastrulation (Wilkins et al. 2008) and newborn mice homozygous for a deletion including *Mtx2* were never scored (binomial $P = 5.7 \times 10^{-6}$; supplementary note 1, Supplementary Material online), we concluded that homozygosity at the goat *POLYCERATE* locus is an early lethal condition.

Remote *Hoxd1* Regulation in Transgenic Mice

These mapping studies identified the *HOXD* gene cluster as being involved in the polycerate phenotype in both sheep and goats. This cluster contains nine homeobox genes encoding transcription factors involved in the organization of the body plan during embryogenesis (Krumlauf 1994). Both their timing of activation and their domains of expression are determined by their respective positions along the gene cluster (Kmita and Duboule 2003). Accordingly, the mouse *Hoxd1* gene is expressed very early on and in the most rostral part of the embryo (fig. 2*a*). In rodents, *Hoxd1* is expressed in crest cell-derived structures (Frohman and Martin 1992), which made this gene a particularly interesting candidate for polyceraty. In addition, a DNA sequence conserved only among pecoran species carrying headgear was identified 15 kb

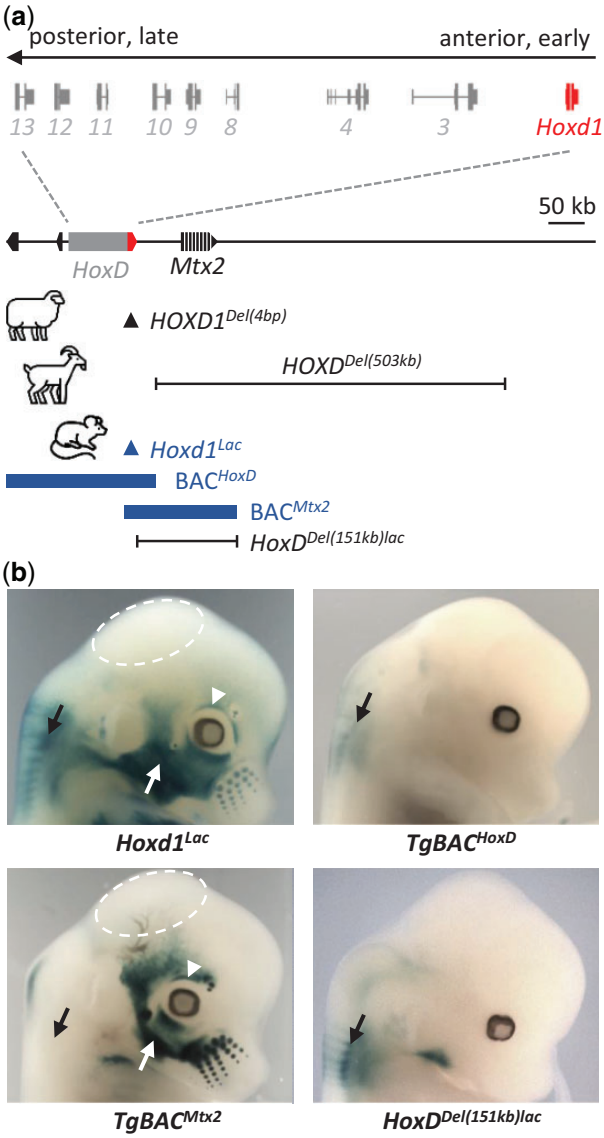


FIG. 2. Regulation of *Hoxd1* expression pattern in crest cell-derived head structures in mouse. (*a*) On top is the structure of the mouse *HoxD* gene cluster with arrows showing the timing and localization of gene expression along the body axis during development. The position of *Hoxd1* is highlighted in red. Below is a 1-Mb view of the locus, with *Hoxd1* in red as well as the relative positions of the *POLYCERATE* variants in sheep (black arrowhead) and goat (black line). Below are depicted the various murine alleles, with the lacZ insertion in *Hoxd1* (blue arrowhead), the two BAC clones (thick blue lines) and the engineered deletion (black line). (*b*) Heads of E12.5–E13.5 mouse fetuses after X-gal staining. The dashed circle highlights the absence of *Hoxd1* expression in the crown (corresponding to the localization of hornbuds in Bovidae), whereas the surrounding dermal cells are positive. The conservation of *Hoxd1* expression in the back of the neck (black arrows) contrasts with the presence/absence of expression in the facial muscle precursors (white arrows) and in the eyelids (arrowhead). The comparison between the four strains indicate that *Hoxd1* expression in all these cranial derivatives is controlled by regulatory elements located in a region orthologous to the proximal portion of the segment deleted in polycerate goats. Sheep, goat and mouse icons were made by "Monkik" from www.thenounproject.com (last accessed November 4, 2020).

downstream of *HOXD1* ("HCE" in Wang et al. [2019]). This sequence, however, is not included in the large indel observed in polycerate goats.

We assessed whether the deletion present in goat may impact the expression of *HOXD1* in cranial crest cells by looking at a series of modified mouse strains either carrying transgenes or where a targeted deletion was induced at the orthologous locus (see Materials and Methods). First, the wide presence of cells expressing *Hoxd1* both in the face and in the cranial derma, the latter being of crest cell origin, was detected in fetuses with a targeted integration of lacZ sequences into the *Hoxd1* gene (fig. 2*b*, $Hoxd1^{Lac}$). Expression was however not scored in the crown region (fig. 2*b*, dashed circle), an area we presumably defined as corresponding to that of horn bud differentiation in Bovidae (Dove 1935; Capitan et al. 2011). Instead, *Hoxd1* was expressed in various amounts in other regions of the head including the eyelids (fig. 2*b*, white arrow and arrowhead), an observation consistent with the abnormal upper eyelids and eyebrows detected in a minority of polycerate sheep and goats (Gascoigne et al. 2017; supplementary figs. 5–7, Supplementary Material online), even though such alterations were not observed in mice lacking *Hoxd1*.

We next tried to localize the underlying regulatory elements by using transgenic BACs with lacZ sequences introduced within *Hoxd1*. A BAC covering the *HoxD* cluster itself did not show any expression in the head, suggesting that regulatory sequences are not located in the gene cluster (fig. 2*b*, $TgBAC^{HoxD}$). In contrast, a transgenic BAC extending in the region upstream of *Hoxd1* and including *Mtx2* gave a subset of the staining observed with $Hoxd1^{Lac}$ (fig. 2*b*, $TgBAC^{Mtx2}$), indicating that some regulatory sequences were located upstream *Hoxd1*, in a region including and surrounding *Mtx2*. The latter result was controlled by using an engineered 151-kb deletion of a largely overlapping region, including a lacZ reporter gene, which expectedly abrogated *Hoxd1* expression in cranial cellular populations (fig. 2*b*, $HoxD^{Del(151\,kb)lac}$). The weaker expression of $Hoxd1^{Lac}$ in the dermal component also seemed to disappear in the latter deletion. As a positive control for the lacZ reporter system, expression of *Hoxd1* in neural derivatives driven by sequences within the *HoxD* cluster was scored, as expected (fig. 2*b*, black arrows). These analyses in mice demonstrated that regulatory sequences driving *Hoxd1* expression in the head are located in a region largely comprised within the deletion determined in goats as causative of polyceraty, further suggesting that the latter deletion abrogates *HOXD1* expression in goat fetuses.

Expression of *HOXD1* in Pecoran Fetuses

To investigate whether the absence of mouse *Hoxd1* expression in the crown region of the head was also observed in pecoran embryos, we isolated heterozygous polycerate and wild-type fetuses both at 70 dpc (days post-coïtum) in goat and at 76 dpc in sheep, two stages where eyelids are fully grown and horn buds can be distinguished (supplementary fig. 8, Supplementary Material online). After microdissection and reverse transcription-quantitative PCR (RT–qPCR), we noticed that in wild-type fetuses of both species, *HOXD1* expression was significantly lower in horn buds than in surrounding tissues (fig. 3*a* and *b*), reminiscent of the weak—if any—expression of *Hoxd1* observed in a comparable region in the mouse. In heterozygous mutant goat fetuses, however, *HOXD1* RNA levels were equally low in all three samples (fig. 3*a*), re-enforcing the idea that the caprine *POLYCERATE* variant negatively affects the expression of *HOXD1*.

In sheep, when primers targeting the second exon of the gene were used (fig. 3*b*, upper histogram and methods), heterozygous mutants for the 4-bp deletion overlapping the splice donor site and control samples displayed similar profiles of RNA expression despite some variation due to slight differences in sampling. However, RT–qPCR with intronic primers revealed significant intron retention in all mutant tissues but horn buds, where expression was likely too low (fig. 3*b* lower histogram). Intron retention is predicted to result in a nonfunctional protein, truncated two residues after the last amino acid encoded by exon 1 and thus lacking the homeodomain, the DNA binding moiety (fig. 3*c* and supplementary fig. 9, Supplementary Material online). Therefore, both *POLYCERATE* variants appear to reduce the amount of functional *HOXD1* RNAs in the horn bud region. We hypothesize that this reduction leads to the extension of the cellular field permissive for horn bud development following the loss of the HOXD1 boundary. This extension may sufficiently elongate the bud region to allow its separation into two distinct organs.

Morphometric Analyses and Topology of the Horn Field

To substantiate this hypothesis, we analyzed variations in horn topology in 61 ovine and 19 caprine skulls from various populations using 3D geometric morphometrics (supplementary table 10, Supplementary Material online). We performed a principal component analysis (PCA) using 16 anatomical landmarks (anatomically homologous) and 100 sliding semilandmarks (geometrically homologous [Bookstein 1992] after scaling and eliminating the effects of translation and rotation thanks to a generalized procrustes analysis [GPA]; Rohlf and Slice 1990). This protocol, using sliding semilandmarks, makes it possible to quantify, visualize, and compare anatomical regions devoid of anatomical landmarks (Gunz and Mitteroecker 2013; see Materials and Methods). We then plotted the first principal components (PCs) to visualize the specimen distribution in the morphospace (fig. 4*a* and supplementary figs. 10–12 and supplementary table 11, Supplementary Material online). The first two axes represented 35.8% and 23.3% of the total shape variability and distinguished the phenotypes and species categories, respectively. Along the first axis, we individualized three subgroups of polycerate specimens in sheep, based on the distances between lateral horns (fig. 4*a*, dlh). Of note, the group displaying the largest dlh (i.e., that with the highest negative values along the *x* axis) had no equivalent in goat, possibly due to early lethal homozygosity (see above).

We looked at the association between genotypes and horn implantation within polycerate animals by measuring the

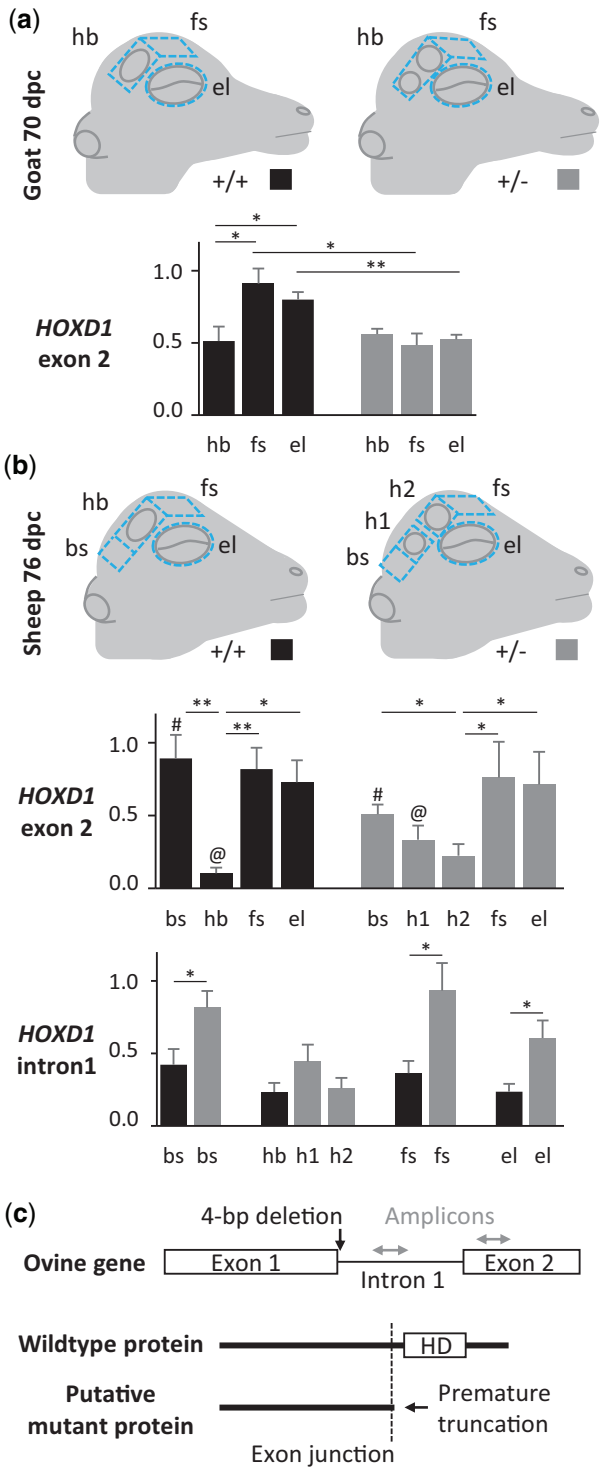


FIG. 3. RT–qPCR gene expression analyses in sheep and goat fetuses. (*a* and *b*) Schemes of the tissues sampled at stage 70 dpc in goat (*a*) and 76 dpc in sheep (*b*) in four control (+/+) and four heterozygous (+/−) polycerate fetuses within each species. bs: skin from the back of the head; hb: skin from the hornbud; h1: skin from the lower horn bud; and h2: skin from the upper horn bud in polycerate specimens; fs: frontal skin; el: eyelids. RT–qPCR gene expression analyses in these tissues are shown below (means and standard errors of the means from triplicate experiments). Gene expression was normalized using *GAPDH*, *H2AFZ*, and *HPRT1* as reference genes. *$P < 0.05$, **$P < 0.01$ (Welch two-sample *t*-test with the alternative hypothesis that the means are not equal). For the sake of clarity, the symbols # and @ were also used to show significant differences ($P < 0.05$) between distant bars. (*c*) Schematic representation of the ovine *HOXD1* gene and corresponding wild-type and putative mutant proteins. The localizations of the amplicons studied in (*b*) are indicated with double arrows. HD: Homeodomain.

distances both between the lateral horns on the left side of the skull (dlhl), and between the upper horns (duh) in 29 rams (supplementary table 12, Supplementary Material online). We found a significant difference in the proportions of homozygous and heterozygous specimens in animals with dlhl≤duh versus dlhl>duh (fig. 4*b*) and no heterozygous animal was found to have dlhl>duh. We computed the theoretical skull shape at the maximum and minimum of PC1 axis (fig. 4*c*) and the corresponding vectors of deformation (fig. 4*d*). The results obtained were consistent with a splitting of horn buds in polycerate animals. This splitting always occurred along the major axis of the ellipse formed by the wild-type horn bud, with an extension of the hemi-horn buds in an area where *HOXD1* expression was detected in wild-type specimens (fig. 4*d* and above). In homozygous animals, the new cellular field was likely larger than in heterozygous, leading to a clearer separation of hemi-horns, whereas heterozygous specimens often displayed partially fused organs, a situation markedly different from the production of additional horns, as observed in subspecies of *Tetracerus quadricornis* (Groves 2003) (supplementary fig. 13, Supplementary Material online).

Conclusions

From these results, we conclude that pecorans have an intrinsic capacity to induce hornbuds within a presumptive head territory. This capacity appears to be associated with the non-expression of the *HOXD1* transcription factor, which is present in surrounding cells and may delimit this field, a function somewhat distinct from the ancestral role of *Hox* genes during development (Krumlauf 1994). Two independent haploinsufficient conditions, in sheep and goat, both involving reduced expression of *HOXD1* presumably lead to the extension of this territory, a condition fully achieved in the complete absence of the wild-type *HOXD1* allele in homozygous polycerate sheep. Although a weak extension of this morphogenetic field may correspond to the growth of twin horns, fused at their bases, a full extension would induce the complete splitting of the horn bud, thus generating a pair of lateral horns. We hypothesize that the initial expression of *HOXD1* in anterior crest cells made this evolutionary co-option possible and thus helped to determine the position and number of horns, which became the distinctive trait of Bovidae.

Materials and Methods

Ethics Statement

All experiments reported in this work comply with the ethical guidelines of both the French National Research Institute for Agriculture, Food and Environment (INRAE) and the University of Geneva, Switzerland. Blood samples were

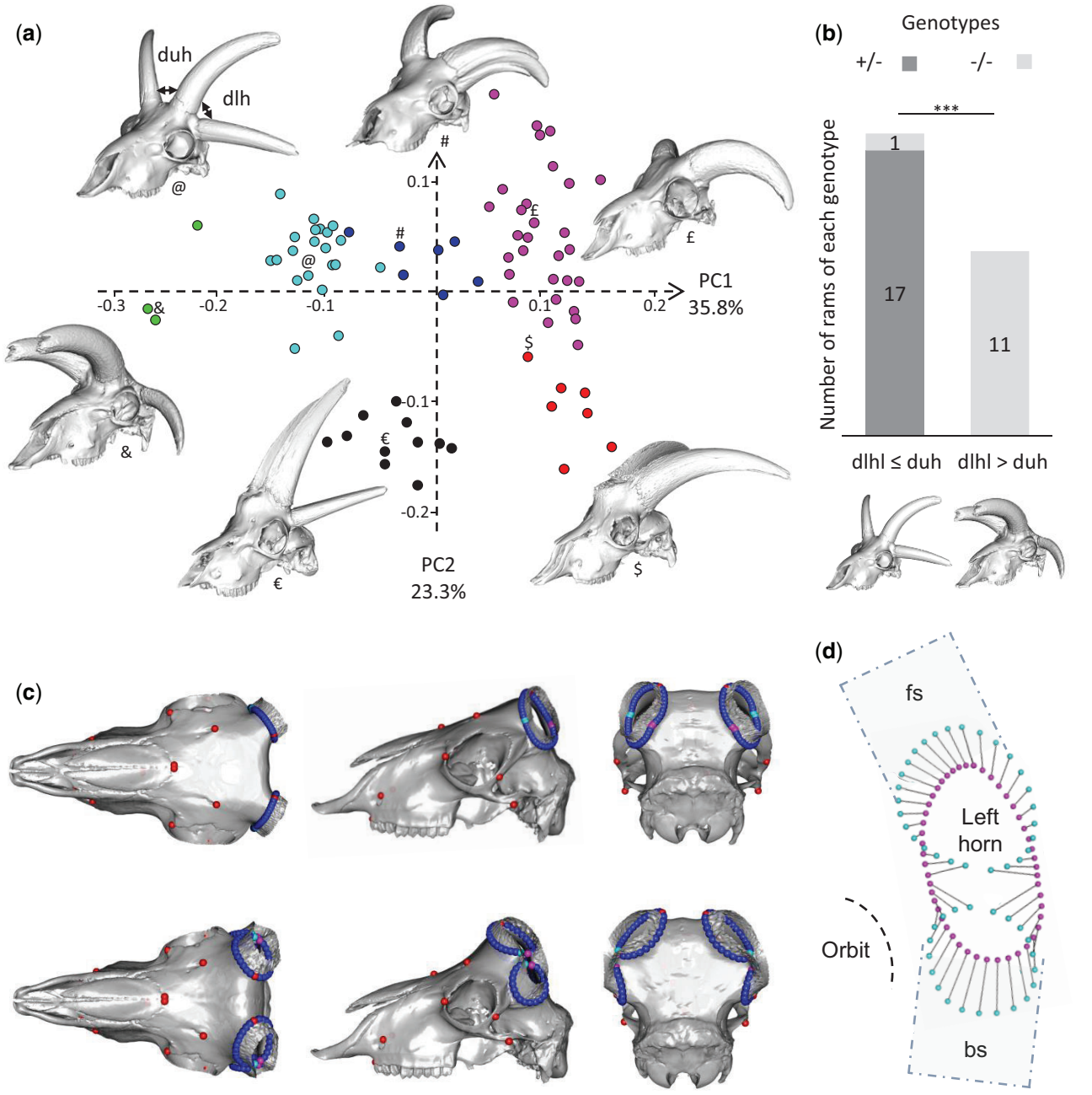


Fig. 4. Results of 3D geometric morphometric analyses of 61 ovine and 19 caprine skulls. (*a*) Distribution of the specimens along the first two axes of the PCA. The proportion of variance explained by the main principal components is indicated on each axis. Green dots: polycerate sheep with a distance between lateral horns (dlh) larger than the distance between upper horns (duh); light blue: polycerate sheep with a dlh≤duh; blue: polycerate sheep with at least two lateral horns partially fused at their basis; purple: wild-type sheep; black: polycerate goats; and red: wild-type goats. Representative specimens illustrate each cluster and symbols are used to indicate their respective locations in the PCA analysis (see supplementary fig. 10, Supplementary Material online, for intraspecies analyses and further information). (*b*) Number of heterozygous (+/−) and homozygous (−/−) polycerate rams among groups of live animals with different dlhl (dlh on the left side) and duh relative sizes (see supplementary table 12, Supplementary Material online, for further information); ***P value $= 3.5 \times 10^{-7}$ (Fisher's exact test). (*c*) Theoretical shapes associated with the maximum (upper three) and minimum values (lower three) of PC1 axis for a sheep skull. Red dots correspond to anatomical landmarks, whereas the other dots correspond to sliding semilandmarks; light blue and purple dots highlight the sites of division of lateral horns. (*d*) Shape differences for the sliding semilandmarks located at the basis of the left horn. Light blue and purple dots correspond to the maximum and minimum values of PC1 axis, respectively. Dashed squares indicate the estimated position of dissected tissues in figure 3 (bs: skin from the back of the head; fs: frontal skin) in which *HOXD1* expression was observed in fetuses.

collected on sheep and goats during routine blood sampling (for annual prophylaxis, paternity testing, or genomic selection purpose) by trained veterinarians and following standard procedures and relevant national guidelines. Sample collection of small ruminants in Switzerland was approved by the Cantonal Committee for Animal Experiments (Canton of Bern; permit 75/16). Ovine and caprine fetuses were produced in an INRAE experimental farm (Bressonvilliers, France) and collected in the INRAE experimental slaughterhouse of Jouy-en-Josas (France). Experiments were performed in strict accordance with the European directive 2010/63/UE and were approved by the local Institutional Animal Care and Use Committee of AgroParisTech/INRAE (COMETHEA, permit number 19/032). All experiments involving mice were performed in agreement with the Swiss law on animal protection (LPA), under license No GE 81/14 (to D.D.). All the samples and data analyzed in the present study were obtained with the permission of breeders, breeding organizations, and research group providers.

Animals

Live Sheep and Goats

Animals from a wide diversity of breeds around the world were involved in at least one of the analyses performed in this study. Briefly, they fall into four categories: 1) Individuals genotyped with Illumina OvineHD or GoatSNP50 (Tosser-Klopp et al. 2014) BeadChip for mapping the *POLYCERATE* locus in both species (supplementary tables 1 and 6, Supplementary Material online). 2) A set of whole-genome sequences used for identifying and filtering candidate mutations (supplementary table 13, Supplementary Material online). 3) Individuals genotyped by PCR and Sanger sequencing for candidate mutations (supplementary tables 3 and 7, Supplementary Material online). 4) Polycerate sheep animals genotyped for verifying putative differences between heterozygous and homozygous individuals in terms of distances between the lateral horns and between the upper horns (supplementary table 12, Supplementary Material online).

Mouse Models

Five different transgenic mouse stocks were used (see supplementary table 14, Supplementary Material online). The $HoxD^{(Del365)}$ allele was produced by CRISPR-Cas9 technology. sgRNA were designed manually, ordered as DNA oligos at Eurogentec, and cloned into px330. sgRNAs were synthetized with HiScribe T7 high yield RNA synthesis kit (New England Biolabs), incubated together with Cas9 mRNA and electroporated into fertilized mouse zygotes (see also supplementary note 1, Supplementary Material online). *The* $HoxD^{(Del151)}$ allele was obtained by using CRE-mediated recombination (Andrey et al. 2013). The Transgenic fetuses from four strains containing different lacZ constructions were collected from stage E12.5 to E.15.5. The $Hoxd1^{Lac}$ strain was obtained by inserting a LacZ cassette in the HindIII site of the second exon of *Hoxd1* (Zakany et al. 2001). The BAC^{HoxD} and BAC^{Mtx2} strains result from the introduction of a LacZ-SV40promoter-*Hoxd1*-zeocin cassette into the HindIII site of the second exon of *Hoxd1* (Schep et al. 2016). The BACs were selected based on their localization on the physical map of the mouse genome (Gregory et al. 2002) and obtained from the RPCI-23 and -24 Mouse (C57BL/6J) BAC Libraries from the Children's Hospital Oakland Research Institute (https://bacpacresources.org/libraries.php, last accessed November 4, 2020). The modified BACs were purified, linearized, and microinjected into mouse fertilized oocytes to obtain each of these strains in a mixed Bl6XCBA hybrid background, by standard procedures. Gene expression analyses were performed on heterozygous specimens. A precise map of the orthologous *Hoxd* region in mouse and goat was obtained by aligning on murine GRCM38/mm10 genome assembly the BAC end sequences and goat genome sequences of 10-kb segments encompassing the breakpoints of the large indel. Alignments were carried out using the BLAT tool from the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgBlat, last accessed November 4, 2020).

Animals Subject to Postmortem Clinical Examination

The eyelids and eyes fundus were examined in a 3-week-old polycerate male Provençale kid who died from a natural cause and a matched control, as well as in an 8-year-old polycerate Jacob ewe and her wild-type half-sister after slaughter.

Ovine and Caprine Fetuses

Fetuses were generated by mating heterozygous polycerate males of the caprine Provençale and ovine Jacob breeds with wild-type cull females after oestrus synchronization. Oestrus cycles were synchronized using intravaginal sponge impregnated with progestagen for 15 days followed by PMSG (Pregnant Mare Serum Gonadotropin) injection 48 h after sponge removal. Pregnant females were anesthetized by electronarcosis and euthanized by immediate exsanguination on day 70 or 76 post-coïtum in the INRAE slaughterhouse of Jouy-en-Josas (France). Directly after, the fetuses were recovered from their genital tracts and exsanguinated. "Skin" samples comprising the epidermis, dermis, and hypodermis were collected at different locations on the left side of the head of the 70 dpc goat and 76 dpc sheep fetuses (see fig. 3) for expression studies. Of note, the skin of the back of the head was sampled slightly more caudally in polycerate animals due to the specific localization of the posterior pair of horns. The same skin samples were collected on the right side of the head with the underlying bone for histological analyses. Four case fetuses and four sex-matched controls were selected in each species for expression studies. Finally, for verification, liver samples were also collected for DNA extraction and subsequent genotyping of the fetuses for the sheep and goat *POLYCERATE* mutations.

Skull Specimens

The skulls from 61 sheep (32 polycerate and 29 wild type) and 19 goats (12 polycerate and 7 wild type) were obtained from different anatomical collections. These specimens were sampled over the last 170 years and originate from a wide variety of populations. Information on horn phenotype, species,

gender, age, population or breed, collection, and year of entry in the collection is given in supplementary table 10, Supplementary Material online.

Phenotyping

The polycerate phenotype is an autosomal dominant trait readily visible on fetuses at 70 dpc in goat and 76 dpc in sheep (supplementary fig. 8, Supplementary Material online). Phenotyping at birth is difficult due to the presence of hairs and it is necessary to wait for after the first month to distinguish horns growing amid fur. In polycerate animals, horns have a nearly circular cross section but, depending on their relative placement, they may progressively fuse at the base with other horns located on the same side of the skull. The growth in width of horns is expected to affect the measure of distances between the lateral horns (dlh) and the upper horns (duh), but not their relative sizes. This, together with the fact that we never observed any case of fusion between the upper horns, led us to consider the dlh/duh ratio on the left side of the head to distinguish different types of four-horned animals in one of the analyses performed in this study. Polyceraty is sometimes associated with defects of the eyelid in both species. Although we did not systematically record this particular phenotype, we performed postmortem clinical examination of the eyelids and eyebrows in one case and one control animal per species (supplementary figs. 5–7, Supplementary Material online).

DNA Extraction

Ovine and caprine DNAs were extracted from hair root, blood, or liver samples using the DNeasy Blood and Tissue Kit (Qiagen). Murine DNA was isolated from ear snip after Proteinase K digestion using standard phenol/chloroform protocol. DNA quality was controlled by electrophoresis and quantified using a Nanodrop spectrophotometer (Thermo Scientific).

IBD-Mapping of Caprine and Ovine *POLYCERATE* Loci

Assuming autosomal dominant inheritance and genetic homogeneity in each of the species investigated, all polycerate animals share at least one copy of the same causative mutation and of a surrounding chromosomal segment inherited-by-descent from a common ancestor. Therefore, comparing SNP array genotyping data of two distantly related polycerate animals is expected to reveal a number of Mendelian incompatibilities (i.e., homozygosity for different alleles) throughout their genomes but not within shared IBD segments. Accordingly, we screened Mendelian incompatibilities in all the possible pair combinations of polycerate × polycerate (4H4H pairs) and polycerate × wild-type (4H2H) individuals. Pairs with a proportion of Medelian incompatibilities below 1 percent of the total number of markers tested were declared as constituted of parent and offspring and were not considered in the analysis. Then, for sliding windows of n markers (n set to 10 in goat and 50 in sheep, considering differences in marker density) we scored the numbers of 4H4H pairs and 4H2H pairs for which "no" versus "at least one" Mendelian inconsistency has been recorded. Finally, we compared the contingency tables produced using Fisher's exact test.

SNP Array Genotypes, Sample, and Variant Pruning

Illumina GoatSNP50 BeadChip genotypes specifically generated for this research and Illumina OvineHD Beadchip genotyping data generated by two previous studies (Greyvenstein et al. 2016; Kijas et al. 2016) were considered in the analyses. Polled (i.e., hornless) animals were removed from the sheep data set. Markers with a minor allele frequency below 5% or which were called in less than 95% of the samples were eliminated. Moreover, in sheep, genotyping data were extracted for markers located in a 10-Mb region (Chr2:127,500,001–138,500,000 on Oar_v4.0 assembly) corresponding approximately to the *HOXD* gene cluster ±5 Mb and encompassing all the mapping intervals of the *POLYCERATE* locus reported in the literature (Greyvenstein et al. 2016; He et al. 2016; Kijas et al. 2016; Ren et al. 2016). The final data sets contained 111 cases, 87 controls, and 2,232 markers in sheep and 35 cases, 51 controls, and 48,345 markers in goat.

Analysis of Whole-Genome Sequences

The genomes of one polycerate Provençale goat and one polycerate Jacob sheep were sequenced specifically for this study. Both were born from polycerate × wild-type crosses and thus were predicted to be heterozygous for the caprine and ovine causative variants, respectively. Paired-end libraries with a 450- (goat) and 235-bp (sheep) insert size were generated using the NEXTflex PCR-Free DNA Sequencing Kit (Bioscientific). Libraries were quantified with the KAPA Library Quantification Kit (Cliniscience), controlled on a High Sensitivity DNA Chip (Agilent), and sequenced on a HiSeq 2500 (with 2 × 100-bp read length in goat) and a HiSeq 3000 (with 2 × 150-bp read length in sheep). The average sequence coverage was 16.7 and 11.1×, for the polycerate goat and sheep individuals, respectively. Additional whole-genome sequences available in public databases were also considered in the analyses. These consisted of FASTQ files (for 10 additional case and 341 control sheep) and of VCF files (for 1,160 goat and 838 sheep control individuals) generated by previous studies (see supplementary table 13, Supplementary Material online). When necessary, the NCBI Genome Remapping Service (https://www.ncbi.nlm.nih.gov/genome/tools/remap, last accessed November 4, 2020) was used to convert positions in VCF files between older and most recent versions of genome assemblies.

The sequence reads from FASTQ files were mapped on goat ARS1 (https://www.ncbi.nlm.nih.gov/assembly/GCF_001704415.1/, last accessed November 4, 2020) and sheep Oar_v4.0 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000298735.2, last accessed November 4, 2020) genome assemblies using the BWA-MEM software v 0.7.17 with default parameters (Li and Durbin 2009) and converted to bam format with v 1.8 of SAMtools (Li et al. 2009). Duplicate reads were marked using Picard tools v 2.18.2 MarkDuplicates option (http://broadinstitute.github.io/picard, last accessed November 4, 2020) and base quality recalibration and indel

realignments were done with v 3.7 of GATK (McKenna et al. 2010). Reads located in the mapping intervals of the ovine and caprine *POLYCERATE* loci ±1 Mb were extracted using SAMtools view option before processing to the calling of SNPs and small indels with GATK-HaplotypeCaller in ERC mode. The minimum read mapping quality and phred-scaled confidence threshold were set to 30 for each sample ("-stand_call_conf 30.0 -mmq 30 -ERC GVCF -variant_index_type LINEAR -variant_index_parameter 128000"). In goats, we retained only heterozygous variants found in the heterozygous polycerate individual and absent from 1,160 control animals, whereas in sheep we focused our attention on variants which were shared (either in heterozygous or homozygous state) in all the 11 polycerate sheep (1 Jacob and 10 Sishui Fur Sheep) and absent from the 1,179 control animals. Finally, to ensure that we did not miss any candidate variants, we performed a detection of structural variants in the same regions using Pindel (Ye et al. 2009) and a visual examination of the whole-genome sequences for 11 goats (1 case, 10 controls) and 22 sheep (11 cases and 11 controls) using IGV (Thorvaldsdóttir et al. 2013). The count command in IGVtools was used to produce ".tdf" files and identify changes in read coverage in the intervals investigated (with parameters: zoom levels = 10, window function = mean, window size = 1,000, and extension factor = 500).

Definition of the Boundaries of the 503-kb Deletion–137-kb Insertion in Goat

The boundaries of variant g.115,652,290_116,155,699delins137kb were reconstructed manually using split read and paired-end read information obtained from IGV. Sequences of reads affected by the mutation were extracted from the .bam file using linux command lines and aligned manually to reconstruct the nucleotide sequence at each fusion point. For verification, amplicons encompassing these fusion points were PCR amplified in a Mastercycler pro thermocycler (Eppendorf) using Go-Taq Flexi DNA Polymerase (Promega), according to the manufacturer's instructions and primers listed in supplementary table 15, Supplementary Material online. Amplicons were purified and bidirectionally sequenced by Eurofins MWG (Hilden, Germany) using conventional Sanger sequencing.

Genotyping of DNA Sequence Variants

SNP and small Indels were genotyped using PCR and Sanger sequencing as described above. PCR primers were designed with Primer3 software (Rozen and Skaletsky 2000) and variants were detected using NovoSNP software (Weckx et al. 2005). Transgene insertions and large indels were genotyped by PCR and electrophoresis on a 2% agarose gel. Ovine variant g.132,832,249_132,832,252del was genotyped with primers TTTGGGGCCACACTAGAATC and CCTAGAGGGGGCCTACGAG, whereas caprine and murine variants were genotyped with the primers listed in supplementary tables 7 and 14, Supplementary Material online, respectively.

Analysis of Nucleotide Sequence Conservation at the *HOXD1* Exon 1–Intron 1 Junction

Nucleotide sequences of the *HOXD1* gene in 103 sarcopterygian and tetrapod species were obtained from the Ensembl (http://www.ensembl.org/index.html, last accessed November 4, 2020; release 98) and UCSC (http://genome.ucsc.edu/, last accessed November 4, 2020) genome browser databases. The localization of the nucleotide sequence (between *MTX2* and *HOXD3*) was verified in each genome assembly to avoid possible confusion with paralogs. In addition, only one sequence was arbitrarily retained when genome assemblies for distinct individuals of the same species were available. Then sequences were put in the same orientation and trimmed to get 40 nucleotides before and 20 nucleotides after the splice donor site of *HOXD1* exon 1. A multispecies alignment was generated with ClustalW software (Thompson et al. 1994), version 2.1 (https://www.genome.jp/tools-bin/clustalw, last accessed November 4, 2020) and a sequence logo was generated using WebLogo (Crooks 2004) (http://weblogo.berkeley.edu/, last accessed November 4, 2020). Information on species, sequence, and genome assemblies are presented in supplementary table 5, Supplementary Material online.

Fluorescence In Situ Hybridization in Goat

Skin biopsies were sampled from one heterozygous polycerate and one wild-type fetuses. Fibroblast cultures and metaphases were obtained according to (Ducos et al. 2000). Nucleotide sequences from the segments of caprine chromosomes 2 and 5 involved in the candidate causative mutation were aligned against bovine bacterial artificial chromosome (BAC) end sequences using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi, last accessed November 4, 2020). Two INRA BAC clones (Eggen et al. 2001) were selected and obtained from the Biological Resources of @BRIDGe facilities (abridge.inrae.fr): INRAb 230B11, targeting the segment deleted on Chr2, and INRAb 348A12, targeting the region of Chr5 that is duplicated and inserted on Chr2. FISH experiments were carried out according to (Yerle et al. 1994). The two BACs were labeled with biotin and digoxygenin, respectively, using the BioPrime DNA Labeling System kit (Life Technologies, Carlsbad, CA). Finally, they were revealed by Alexa 594 conjugated to streptavidin (Molecular Probes, Eugene, OR) and FITC conjugated mouse antidigoxygenin antibodies (Sigma, St Louis, MO).

Histological Analyses

Tissues were fixed in paraformaldehyde (4%) for 24 h at +4°C. Samples were subsequently dehydrated in a graded ethanol series, cleared with xylene, and embedded in paraffin wax. Microtome sections (5 µm, Leica RM2245) were mounted on adhesive slides (Klinipath-KP-PRINTER ADHESIVES), deparaffinized, and stained with hematoxylin, eosin, and saffron (HES). Slides were scanned with the Pannoramic Scan 150 (3D Histech) and analyzed with the CaseCenter 2.9 viewer (3D Histech).

Quantitative RT–PCR

RNA was extracted using the RNeasy Mini Kit (Qiagen). Super-Script II (Invitrogen) was used to synthesize cDNA from 2 µg of total RNA isolated from each tissue sampled in 70 dpc goat and 76 dpc sheep fetuses. Gene sequences were obtained from Ensembl v92 (www.ensembl.org, last accessed November 4, 2020) and PCR primers (supplementary table 16, Supplementary Material online) were designed using Primer Express Software for Real-Time PCR 3.0 (Applied Biosystems). Primer efficiency and specificity were evaluated on genomic DNA in each species. Quantitative PCR was performed in triplicate with 2 ng of cDNA using the Absolute Blue SYBR Green ROX mix (Thermo Fisher Scientific) and the StepOnePlus Real-Time PCR System (Applied Biosystems). The expression stability of five genes (*RPLP0*, *GAPDH*, *H2AFZ*, *YWHAZ*, and *HPRT1*) was tested at each time point using the GeNorm program (Vandesompele et al. 2002) to identify appropriate qRT–PCR normalizing genes. Three normalizing genes (*GAPDH*, *H2AFZ*, and *HPRT1*) were retained and the results were analyzed with qBase software (Hellemans et al. 2007).

Consequences of Intron Retention Due to the Four-Nucleotide Deletion in *HOXD1* Intron 1

The complete nucleotide sequence of ovine *HOXD1* gene was obtained from Ensembl v97. A mutant mRNA characterized by 1) a retention of intron 1 and 2) a deletion of nucleotides located at position +4 to +7 bp after the end of exon 1 was designed. This mutant mRNA was translated using ExPASy Translate tool (https://web.expasy.org/translate/, last accessed November 4, 2020). Information on HOXD1 functional domains was obtained from UniProt Knowledgebase (https://www.uniprot.org/uniprot/W5Q7P8, last accessed November 4, 2020).

3D Geometric Morphometrics

3D models were generated for 80 skulls consisting of 32 polycerate and 29 wild-type sheep specimens, as well as 12 polycerate and 7 wild-type goat specimens (for information on skulls and reconstruction methods, see supplementary table 10, Supplementary Material online). Most of the 3D models (n = 47) were reconstructed using a Breuckmann StereoScan structured light scanner and its dedicated software OptoCat (AICON 3D systems, Meersburg, Germany). Twenty-nine skulls were digitized with the Artec Eva structured-light scanner and ScanStudioHD software v12.1.1.12 (Artec 3D, Luxembourg, Luxembourg). In addition, four skulls were digitized with a photogrammetric approach, similar to that described in (Evin et al. 2016). In brief, hundred pictures per sample were taken on different angles and inclinations with a Nikon D3300 camera equipped with an AF-S Micro Nikkor 85 mm lens (Nikon, Tokyo, Japan) and a self-made fully automatic turntable. Then 3D models were reconstructed with the ReCap Photo software (Autodesk, San Rafael, CA). Previous studies indicated no significant differences between 3D models obtained with 3D scanners or photogrammetry (Evin et al. 2016; Fau et al. 2016). Both approaches are comparable in terms of measurement error (less than 1 mm). Bone surfaces were extracted as meshes and geometric inconsistencies (i.e., noise, holes) were cleaned using Geomagic software (3D Systems, Rock Hill).

For shape analyses, 116 3D landmarks and sliding semilandmarks were placed on each specimen by the same operator using the IDAV Landmark software (Wiley et al. 2005) v3.0. Out of them, 16 were anatomical landmarks, and 100 were sliding semilandmarks individually placed around the basis of the horns on the suture between the bony core and the frontal bone. On each side, the first of these 50 sliding semilandmarks was placed on the upper horn, at the intersection between the upper ridge of the bony core and the suture previously mentioned. Details on landmark locations on polycerate and wild-type specimens are provided in supplementary table 11 and supplementary figure 11, Supplementary Material online.

Following the procedure detailed by (Botton-Divet et al. 2015), a template was created using the specimen 2000-438 on which all anatomical landmarks and surface sliding semilandmarks were placed. Then, a semiautomatic point placement was performed (Gunz and Mitteroecker 2013) to project sliding semilandmarks on the surface of the other 3D digitized skulls. Sliding semilandmarks on surfaces and curves were allowed to slide in order to minimize the bending energy of a thin plate spline (TPS) between each 3D meshes and the template. After this first TPS relaxation using the template, three iterative relaxations were performed using the Procrustes consensus of the previous step as a reference.

To remove nonshape variation (i.e., differences in position, scale, and orientation of the configurations) and provide optimal comparability between the specimens, we performed a GPA (Rohlf and Slice 1990). Since our data set contained more variables than observations, we performed a PCA on the procrustes residuals to reduce dimensionality, as recommended by (Gunz and Mitteroecker 2013), and plotted the first Principal Components (PCs) to visualize the specimen distribution in the morphospace. In addition, the mean shape of our sample was used to compute theoretical shapes associated with the maximum and minimum of both sides of the first PC axis for each species using thin plate spline. GPA, PCA, and shape computations were done using the "Morpho" and "geomorph" packages (Adams and Otárola-Castillo 2013; Adams et al. 2018; Schlager 2018) in the R environment (R Core Team 2018).

Repeatability and Reproducibility of Landmark Placement

The 116 landmarks and sliding semilandmarks were placed ten times independently on the skulls from two polycerate and two control male sheep sampled between 1852 and 1909 in Tunisia (A-12130, A12132, 1909-4) and neighboring Algeria (A12157; see supplementary table 10, Supplementary Material online). The measurements were superimposed using a GPA and analyzed using a PCA. Since the variation within specimens was clearly smaller than the variation between specimens (supplementary fig. 12, Supplementary Material online), we considered that the 116 landmarks and sliding semilandmarks were precise enough to describe shape variation.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank L. Orlando (Universities of Toulouse III, France and Copenhagen, Denmark) for tentatively extracting DNA from museum skull specimens as well as C. Hozé, F. Lejuste, and A Michenet (ALLICE), R. McCulloch (CSIRO), S. Chahory, and C. Degueurce (ENVA), F. Andreoletti, M. Boussaha, J. Kergosien, D. Mauchand, M. Femenia, N. Perrot, and M. Vilotte (INRAE), B. Camus-Allanic (LABOGENA DNA), J. Peters (LMU), C. Bens, A. Delapré, and A.Verguin (MNHN), L. Ludes-Fraulob (MZS), and B. Mascrez (University of Geneva) for their assistance. We also thank the staff of the INRAE experimental unit UE 1298 SAAJ for animal husbandry and management, as well as breeders and zoological parcs for making animals available for sampling and for providing pictures. Contributors include in particular the Capgènes breeding company (France), L. Pachot (Mouton Village, Vasles, France), A. Archiloque (France), L. Fiorenzi (Az. Agr. Madonna delle Alpi, Italy), the Schafzuchtverein Jakobschaf Schweiz (Switzerland), Tierpark Hamm (Germany), A. Schumann (Germany), and Dr A. Ennaifer (Zoological Park of Tunis, Tunisia). Finally, the authors thank the VarGoats Consortium for allowing variant filtering against their data set. The Vargoats project was supported by France Génomique (Grant No. ANR-10-INBS-0009). This work was supported by APIS-GENE (Grant AKELOS) and the Swiss National Research Foundation (Grant No. 310030B_138662 to D.D.). A.Hi. was supported by a PhD fellowship from the University of Geneva.

Author Contributions

O.G., D.R., T.H., N.C., C.M.R., B.J.H., and J.K. provided Illumina OvineHD Beadchip genotyping data and related phenotypes. A.C. mapped the ovine and caprine polycerate loci. C.Dr., C.D.-B., D.B., I.M, L.P., O.G., T.H., G.B., F.M., N.H., J.P., S.B.J., J.H., R.R., I.P., J.A.L., L.G., D.R., E.V.M.-K., N.C., B.J.H, J.K., and G.T.-K. provided samples and phenotypes. D.E. and C.Do. performed whole-genome sequencing from one polycerate Provençale goat and one polycerate Jacob sheep. G.T.-K. provided control whole-genome sequences from sheep and goats. A.C., P.B., and M.N.-S. performed variant calling, annotation, and screening for candidate variants. A.C. and A.A.-B. analyzed sequence conservation and annotated the gene content of the mapping intervals. M.-C.D., C.Gr., and A.Hi. extracted DNA. M.-C.D., A.C., C.Gr., and A.Hi. performed PCR for Sanger sequencing and for genotyping by PCR and electrophoresis or PCR and Sanger sequencing. A.P. performed FISH analysis. A.Hi., J.Z., and D.D. produced and studied mouse models. E.P. provided access to laboratory and experimental farm facilities. A.C., A.A.-B., M.-C.D., C.Gr., and E.P. sampled ovine and caprine fetuses. A.A.-B. and M.-C.D. extracted RNA, performed qRT–PCR, and analyzed the results. A.Bo., J.R., A.C., A.A.-B., and M.-C.D. performed histological analyses. A.C., M.S., and A.Ha. performed 3D data acquisition of skulls. A.Bl. provided access to a light scanner for 3D data acquisition. O.P., J.L., R.S., M.-D.W., R.-M.A., and C.Gu. provided access to skull specimens and related information. A.C. performed morphometric analyses. R.C. provided software and expertise in morphometric analyses. A.C. (Bovidae) and D.D. (mouse) designed the studies and wrote the manuscript, which was accepted or revised by all authors.

Data Availability

Raw sequencing data reported in this study were deposited to the European Variation Archive (EVA, https://www.ebi.ac.uk/eva/) under accession number PRJEB39341. Sequences from previous studies can be found at the following URL (www.goatgenome.org/vargoats_data_access.htm) or in the NCBI BioProject and EVA databases under accession numbers PRJEB6025, PRJEB6495, PRJEB9911, PRJEB14098, PRJEB14418, PRJEB15642, PRJEB23437, PRJEB31241, PRJEB31930, PRJEB32110, PRJEB35553, PRJEB35682, PRJEB37460, PRJEB39341, PRJEB39341, and PRJNA624020. Illumina GoatSNP50 Beadchip genotyping data generated for this study have been deposited in the Dryad Digital Repository (doi: 10.5061/dryad.rxwdbrv6n). Illumina OvineHD Beadchip genotyping data from previous studies can be found in the same repository (doi: 10.5061/dryad.6t34b and 10.5061/dryad.1p7sf). Coordinates of landmarks and source data underlying figures 3 and 4, and supplementary figures 2, 10, and 11, Supplementary Material online, are provided as a Source Data file.

References

Adams DC, Collyer M, Kaliontzopoulou A. 2018. Geometric morphometric analyses of 2D/3D landmark data. Available from: http://kambing.ui.ac.id/cran/web/packages/geomorph/geomorph.pdf.

Adams DC, Otárola-Castillo E. 2013. geomorph: an r package for the collection and analysis of geometric morphometric shape data. *Methods Ecol Evol*. 4(4):393–399.

Allais-Bonnet A, Grohs C, Medugorac I, Krebs S, Djari A, Graf A, Fritz S, Seichter D, Baur A, Russ I, et al. 2013. Novel insights into the bovine polled phenotype and horn ontogenesis in Bovidae. *PLoS One* 8(5):e63512.

Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu M, Trono D, Spitz F, Duboule D. 2013. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* 340(6137):1234167.

Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, Lee J, Lam ET, Liachko I, Sullivan ST, et al. 2017. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat Genet* . 49(4):643–650.

Bookstein FL. 1992. Morphometric tools for landmark data: geometry and biology. 1st ed. Cambridge, UK: Cambridge University Press. Available from: https://www.cambridge.org/core/product/identifier/9780511573064/type/book.

Botton-Divet L, Houssaye A, Herrel A, Fabre A-C, Cornette R. 2015. Tools for quantitative form description; an evaluation of different software packages for semi-landmark analysis. *PeerJ* 3:e1417.

Capitan A, Allais-Bonnet A, Pinton A, Marquant-Le Guienne B, Le Bourhis D, Grohs C, Bouet S, Clément L, Salas-Cortes L, Venot E, et al. 2012. A 3.7 Mb deletion encompassing ZEB2 causes a novel polled and multisystemic syndrome in the progeny of a somatic mosaic bull. *PLoS One* 7(11):e49084.

Capitan A, Grohs C, Weiss B, Rossignol M-N, Reversé P, Eggen A. 2011. A newly described bovine type 2 scurs syndrome segregates with a frame-shift mutation in TWIST1. PLoS One 6:e22242.

Crooks GE. 2004. WebLogo: a sequence logo generator. *Genome Res*. 14(6):1188–1190.
Davis EB, Brakora KA, Lee AH. 2011. Evolution of ruminant headgear: a review. *Proc R Soc B*. 278(1720):2857–2865.
Dove WF. 1935. The physiology of horn growth: a study of the morphogenesis, the interaction of tissues and the evolutionary processes of a Mendelian recessive character by means of transplantation of tissues. *J Exp Zool*. 69(3):347–405.
Ducos A, Dumont P, Séguéla A, Pinton A, Berland H, Brun-Baronnat C, Darré A, Guienne B, Humblot P, Boichard D, et al. 2000. A new reciprocal translocation in a subfertile bull. *Genet Sel Evol*. 32(6):589–598.
Durkin K, Coppieters W, Drögemüller C, Ahariz N, Cambisano N, Druet T, Fasquelle C, Haile A, Horin P, Huang L, et al. 2012. Serial translocation by means of circular intermediates underlies colour sidedness in cattle. *Nature* 482(7383):81–84.
Eggen A, Gautier M, Billaut A, Petit E, Hayes H, Laurent P, Urban C, Pfister-Genskow M, Eilertsen K, Bishop MD. 2001. Construction and characterization of a bovine BAC library with four genome-equivalent coverage. *Genet Sel Evol*. 33(5):543–548.
Epstein H. 1971. *The origin of the domestic animals of Africa*. Vol. 2. New York: Africana Publishing Corporation.
Evin A, Souter T, Hulme-Beaman A, Ameen C, Allen R, Viacava P, Larson G, Cucchi T, Dobney K. 2016. The use of close-range photogrammetry in zooarchaeology: creating accurate 3D models of wolf crania to study dog domestication. *J Archaeol Sci*. 9:87–93.
Fau M, Cornette R, Houssaye A. 2016. Photogrammetry for 3D digitizing bones of mounted skeletons: potential and limits. *Comptes Rendus Palevol*. 15(8):968–977.
Frohman MA, Martin GR. 1992. Isolation and analysis of embryonic expression of Hox-4.9, a member of the murine labial-like gene family. *Mech Dev*. 38(1):55–67.
Gascoigne E, Williams DL, Reyher KK. 2017. Survey of prevalence and investigation of predictors and staining patterns of the split upper eyelid defect in Hebridean sheep. *Veterinary Record*. 181(7):167–167.
Gregory SG, Sekhon M, Schein J, Zhao S, Osoegawa K, Scott CE, Evans RS, Burridge PW, Cox TV, Fox CA, et al. 2002. A physical map of the mouse genome. *Nature* 418(6899):743–750.
Greyvenstein OFC, Reich CM, van Marle-Koster E, Riley DG, Hayes BJ. 2016. Polyceraty (multi-horns) in Damara sheep maps to ovine chromosome 2. *Anim Genet*. 47(2):263–266.
Groves C. 2003. Taxonomy of ungulates of the Indian subcontinent. *J Bombay Nat Hist Soc*. 100:314–362.
Gunz P, Mitteroecker P. 2013. Semilandmarks: a method for quantifying curves and surfaces. *Hystrix It J Mamm*. 24(1):103–109.
He X, Zhou Z, Pu Y, Chen X, Ma Y, Jiang L. 2016. Mapping the four-horned locus and testing the polled locus in three Chinese sheep breeds. *Anim Genet*. 47(5):623–627.
Heitzmann A. 2006. Trianon. La ferme du Hameau. *Versalia* 9(1):114–129.
Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol*. 8(2):R19.
Kijas JW, Hadfield T, Naval Sanchez M, Cockett N. 2016. Genome-wide association reveals the locus responsible for four-horned ruminant. *Anim Genet*. 47(2):258–262.
Kmita M, Duboule D. 2003. Organizing axes in time and space; 25 years of colinear tinkering. *Science* 301(5631):331–333.
Krumlauf R. 1994. Hox genes in vertebrate development. *Cell* 78(2):191–201.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760. *England)*
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
Lincoln GA. 1973. Appearance of antler pedicles in early foetal life in red deer. *J Embryol Exp Morphol*. 29:431–437.
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 20(9):1297–1303.
Putelat O. 2005. Le bestiaire polycère. *Revue De Paléobiologie*. V. Spéc. 10:293–301.
R Core Team. 2018. R: a language and environment for statistical computing. Vienne: R Foundation for Statistical Computing. Available from: https://www.R-project.org/
Ren X, Yang G-L, Peng W-F, Zhao Y-X, Zhang M, Chen Z-H, Wu F-A, Kantanen J, Shen M, Li M-H. 2016. A genome-wide association study identifies a genomic region for the polycerate phenotype in sheep. *Sci Rep*. 6(1):21111.
Rohlf FJ, Slice D. 1990. Extensions of the procrustes method for the optimal superimposition of landmarks. *Syst Zool*. 39(1):40.
Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*. 132:365–386.
Schep R, Necsulea A, Rodríguez-Carballo E, Guerreiro I, Andrey G, Nguyen Huynh TH, Marcet V, Zákány J, Duboule D, Beccari L. 2016. Control of Hoxd gene transcription in the mammary bud by hijacking a preexisting regulatory landscape. *Proc Natl Acad Sci U S A*. 113(48):E7720–E7729.
Schlager S. 2018. Morpho: calculations and vizualisations related to geometric morphometrics. Available from: https://cran.r-project.org/web/packages/Morpho/Morpho.pdf.
Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22(22):4673–4680.
Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinformatics*. 14(2):178–192.
Tosser-Klopp G, Bardou P, Bouchez O, Cabau C, Crooijmans R, Dong Y, Donnadieu-Tonon C, Eggen A, Heuven HCM, Jamli S. 2014. Design and characterization of a 52K SNP chip for goats. *PLoS One* 9:e86227.
Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*. 3(7):research0034.1.
Wang Y, Zhang C, Wang N, Li Z, Heller R, Liu R, Zhao Y, Han J, Pan X, Zheng Z, et al. 2019. Genetic basis of ruminant headgear and rapid antler regeneration. *Science* 364(6446):eaav6335.
Weckx S, Del FJ, Rademakers R, Claes L, Cruts M, De Jonghe P, Van Broeckhoven C, De Rijk P. 2005. novoSNP, a novel computational tool for sequence variation discovery. *Genome Res*. 15(3):436–442.
Wiley DF, Amenta N, Alcantara DA, Ghosh D, Kil YJ, Delson E, Harcourt-Smith W, Rohlf FJ, St. John K, Hamann B. 2005. Evolutionary morphing. In: VIS 05. IEEE Visualization, 2005. Minneapolis (MN): IEEE. p. 431–438. Available from: http://ieeexplore.ieee.org/document/1532826/
Wilkins SJ, Yoong S, Verkade H, Mizoguchi T, Plowman SJ, Hancock JF, Kikuchi Y, Heath JK, Perkins AC. 2008. Mtx2 directs zebrafish morphogenetic movements during epiboly by regulating microfilament formation. *Dev Biol*. 314(1):12–22.
Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 25(21):2865–2871.
Yerle M, Goureau A, Gellin J, Le Tissier P, Moran C. 1994. Rapid mapping of cosmid clones on pig chromosomes by fluorescence in situ hybridization. *Mammal Genome*. 5(1):34–37.
Zakany J, Kmita M, Alarcon P, de la Pompa JL, Duboule D. 2001. Localized and transient transcription of Hox genes suggests a link between patterning and the segmentation clock. *Cell* 106(2):207–217.
Zhang M. 1998. Statistical features of human exons and their flanking regions. *Hum Mol Genet*. 7(5):919–932.

BIBLIOGRAPHY

Acemel RD, Maeso I, Gómez-Skarmeta JL. 2017. Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals. *WIREs Dev Biol* **6**: e265. https://onlinelibrary.wiley.com/doi/abs/10.1002/wdev.265 (Accessed April 20, 2020).

Acemel RD, Tena JJ, Irastorza-Azcarate I, Marletaz F, Gomez-Marin C, de la Calle-Mustienes E, Bertrand S, Diaz SG, Aldea D, Aury J-M, et al. 2016. A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. *Nat Genet* **48**: 336–341. http://dx.doi.org/10.1038/ng.3497.

Allais-Bonnet A, Hintermann A, Deloche M-C, Cornette R, Bardou P, Naval-Sanchez M, Pinton A, Haruda A, Grohs C, Zakany J, et al. 2021. Analysis of Polycerate Mutants Reveals the Evolutionary Co-option of HOXD1 for Horn Patterning in Bovidae. *Mol Biol Evol*. https://doi.org/10.1093/molbev/msab021 (Accessed April 22, 2021).

Amemiya CT, Prohaska SJ, Hill-Force A, Cook A, Wasserscheid J, Ferrier DEK, Pascual-Anaya J, Garcia-Fernàndez J, Dewar K, Stadler PF. 2008. The amphioxus Hox cluster: characterization, comparative genomics, and evolution. *J Exp Zoolog B Mol Dev Evol* **310B**: 465–477. http://onlinelibrary.wiley.com/doi/10.1002/jez.b.21213/abstract (Accessed May 20, 2016).

Anania C, Lupiáñez DG. 2020. Order and disorder: abnormal 3D chromatin organization in human disease. *Brief Funct Genomics* **19**: 128–138.

Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu M, Trono D, Spitz F, Duboule D. 2013. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* **340**: 1234167.

Bachiller D, Macías A, Duboule D, Morata G. 1994. Conservation of a functional hierarchy between mammalian and insect Hox/HOM genes. *EMBO J* **13**: 1930–1941.

Banerji J, Rusconi S, Schaffner W. 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**: 299–308.

Bender W, Akam M, Karch F, Beachy PA, Peifer M, Spierer P, Lewis EB, Hogness DS. 1983. Molecular Genetics of the Bithorax Complex in Drosophila melanogaster. *Science* **221**: 23–29.

Berlivet S, Paquette D, Dumouchel A, Langlais D, Dostie J, Kmita M. 2013. Clustering of Tissue-Specific Sub-TADs Accompanies the Regulation of HoxA Genes in Developing Limbs. *PLOS Genet* **9**: e1004018. https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004018 (Accessed April 26, 2020).

Bolt CC, Duboule D. 2020. The regulatory landscapes of developmental genes. *Development* **147**: dev171736. http://dev.biologists.org/lookup/doi/10.1242/dev.171736 (Accessed April 19, 2020).

Bolt CC, Lopez-Delisle L, Mascrez B, Duboule D. 2021. MESOMELIC DYSPLASIAS ASSOCIATED WITH THE HOXD LOCUS ARE CAUSED BY REGULATORY REALLOCATIONS. *bioRxiv* 2021.02.01.429171. https://www.biorxiv.org/content/10.1101/2021.02.01.429171v1 (Accessed April 22, 2021).

Briggs DEG. 2015. The Cambrian explosion. *Curr Biol* **25**: R864–R868. https://www.cell.com/current-biology/abstract/S0960-9822(15)00498-4 (Accessed February 10, 2021).

Burke AC. 2007. Development and evolution of the vertebrate mesoderm. *Dev Dyn* **236**: 2369–2370. https://anatomypubs.onlinelibrary.wiley.com/doi/abs/10.1002/dvdy.21290 (Accessed August 13, 2021).

Cardozo Gizzi AM, Cattoni DI, Nollmann M. 2020. TADs or no TADS: Lessons From Single-cell Imaging of Chromosome Architecture. *J Mol Biol* **432**: 682–693. https://www.sciencedirect.com/science/article/pii/S0022283619307442 (Accessed June 20, 2021).

Carroll SB. 2008. Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. *Cell* **134**: 25–36. http://www.sciencedirect.com/science/article/pii/S0092867408008179 (Accessed May 22, 2016).

Clément Y, Torbey P, Gilardi-Hebenstreit P, Crollius HR. 2020. Enhancer–gene maps in the human and zebrafish genomes using evolutionary linkage conservation. *Nucleic Acids Res* **48**: 2357–2371. https://doi.org/10.1093/nar/gkz1199 (Accessed June 25, 2021).

Darbellay F, Bochaton C, Lopez-Delisle L, Mascrez B, Tschopp P, Delpretti S, Zakany J, Duboule D. 2019. The constrained architecture of mammalian Hox gene clusters. *Proc Natl Acad Sci U S A* **116**: 13424–13433. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6613147/ (Accessed May 3, 2020).

Darbellay F, Duboule D. 2016. Chapter Sixteen - Topological Domains, Metagenes, and the Emergence of Pleiotropic Regulations at Hox Loci. In *Current Topics in Developmental Biology* (ed. P.M. Wassarman), Vol. 116 of *Essays on Developmental Biology, Part A*, pp. 299–314, Academic Press http://www.sciencedirect.com/science/article/pii/S007021531500188X (Accessed April 26, 2016).

de Laat W, Duboule D. 2013. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**: 499–506. http://www.nature.com/nature/journal/v502/n7472/full/nature12753.html (Accessed May 20, 2015).

Dehal P, Boore JL. 2005. Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate. *PLOS Biol* **3**: e314. https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0030314 (Accessed February 16, 2021).

Delpretti S, Montavon T, Leleu M, Joye E, Tzika A, Milinkovitch M, Duboule D. 2013. Multiple Enhancers Regulate Hoxd Genes and the Hotdog LncRNA during Cecum Budding. *Cell Rep* **5**: 137–150. http://www.sciencedirect.com/science/article/pii/S2211124713005007 (Accessed May 23, 2016).

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature* **485**: 376–380. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3356448/.

Dollé P, Izpisúa-Belmonte JC, Brown JM, Tickle C, Duboule D. 1991. HOX-4 genes and the morphogenesis of mammalian genitalia. *Genes Dev* **5**: 1767–1776. http://genesdev.cshlp.org/content/5/10/1767 (Accessed May 29, 2016).

Dollé P, Izpisúa-Belmonte JC, Falkenstein H, Renucci A, Duboule D. 1989. Coordinate expression of the murine Hox-5 complex homoeobox-containing genes during limb pattern formation. *Nature* **342**: 767–772.

Du G, Li H, Ding Y, Jiang S, Hong H, Gan J, Wang L, Yang Y, Li Y, Huang X, et al. 2021. The hierarchical folding dynamics of topologically associating domains are closely

related to transcriptional abnormalities in cancers. *Comput Struct Biotechnol J* **19**: 1684–1693.

Duboule. 1996. Teleost HoxD and HoxA genes: comparison with tetrapods and functional evolution of the HOXD complex. *Mech Dev* **54**: 9–21. https://www.sciencedirect.com/science/article/pii/0925477395004556 (Accessed June 4, 2021).

Duboule D. 1991a. Patterning in the vertebrate limb. *Curr Opin Genet Dev* **1**: 211–216. http://www.sciencedirect.com/science/article/pii/S0959437X05800723 (Accessed May 30, 2016).

Duboule D. 2007. The rise and fall of Hox gene clusters. *Development* **134**: 2549–2560. http://dev.biologists.org/content/134/14/2549 (Accessed May 29, 2015).

Duboule D, Baron A, Mähl P, Galliot B. 1986. A new homeo-box is present in overlapping cosmid clones which define the mouse Hox-1 locus. *EMBO J* **5**: 1973–1980. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1167066/ (Accessed May 8, 2015).

Duboule D, Dollé P. 1989. The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. *EMBO J* **8**: 1497–1505. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC400980/ (Accessed April 18, 2015).

Duboule D, Wilkins AS. 1998. The evolution of `bricolage'. *Trends Genet* **14**: 54–59. https://www.sciencedirect.com/science/article/pii/S0168952597013589 (Accessed June 4, 2021).

Duboule J-CI-B. 1991b. Expression of the homeobox Hox-4 genes and the specification of position in chick wing development | Nature. https://www.nature.com/articles/350585a0 (Accessed June 2, 2021).

Edwards SL, Beesley J, French JD, Dunning AM. 2013. Beyond GWASs: Illuminating the Dark Road from Association to Function. *Am J Hum Genet* **93**: 779–797. https://www.sciencedirect.com/science/article/pii/S0002929713004692 (Accessed June 25, 2021).

Finn EH, Misteli T. 2019. Molecular basis and biological function of variability in spatial genome organization. *Science* **365**: eaaw9498. https://science.sciencemag.org/content/365/6457/eaaw9498 (Accessed September 13, 2019).

Fjose A, McGinnis WJ, Gehring WJ. 1985. Isolation of a homoeo box-containing gene from the engrailed region of Drosophila and the spatial distribution of its transcripts.

Nature **313**: 284–289. https://www.nature.com/articles/313284a0 (Accessed June 16, 2021).

Furlong EEM, Levine M. 2018. Developmental enhancers and chromosome topology. *Science* **361**: 1341–1345. http://science.sciencemag.org/content/361/6409/1341 (Accessed March 6, 2019).

Galliot B, de Vargas C, Miller D. 1999. Evolution of homeobox genes: Q 50 Paired-like genes founded the Paired class. *Dev Genes Evol* **209**: 186–197. http://link.springer.com/10.1007/s004270050243 (Accessed June 16, 2021).

Garber RL, Kuroiwa A, Gehring WJ. 1983. Genomic and cDNA clones of the homeotic locus Antennapedia in Drosophila. *EMBO J* **2**: 2027–2036.

Garcia-Fernàndez J. 2005. The genesis and evolution of homeobox gene clusters. *Nat Rev Genet* **6**: 881–892. http://www.nature.com/nrg/journal/v6/n12/full/nrg1723.html (Accessed March 17, 2015).

Garcia-Fernàndez J, Holland PWH. 1994. Archetypal organization of the amphioxus Hox gene cluster. *Nature* **370**: 563–566. https://www.nature.com/articles/370563a0 (Accessed April 28, 2020).

Gaunt, Duboule. 1988. Spatially restricted domains of homeo-gene transcripts in mouse embryos: relation to a segmented body plan.

Gaunt SJ, Miller JR, Powell DJ, Duboule D. 1986. Homoeobox gene expression in mouse embryos varies with position by the primitive streak stage. *Nature* **324**: 662–664.

Gentile C, Berlivet S, Mayran A, Paquette D, Guerard-Millet F, Bajon E, Dostie J, Kmita M. 2019. PRC2-Associated Chromatin Contacts in the Developing Limb Reveal a Possible Mechanism for the Atypical Role of PRC2 in HoxA Gene Expression. *Dev Cell* **50**: 184-196.e4. https://www.cell.com/developmental-cell/abstract/S1534-5807(19)30419-8 (Accessed August 9, 2021).

Gérard M, Zákány J, Duboule D. 1997. Interspecies exchange of a Hoxd enhancer in vivo induces premature transcription and anterior shift of the sacrum. *Dev Biol* **190**: 32–40.

Ghavi-Helm Y. 2020. Functional Consequences of Chromosomal Rearrangements on Gene Expression: Not So Deleterious After All? *J Mol Biol* **432**: 665–675. http://www.sciencedirect.com/science/article/pii/S0022283619305649 (Accessed February 2, 2021).

Gonzalez F, Duboule D, Spitz F. 2007. Transgenic analysis of Hoxd gene regulation during digit development. *Dev Biol* **306**: 847–859.

http://www.sciencedirect.com/science/article/pii/S0012160607002278 (Accessed May 23, 2016).

Guerreiro I, Gitto S, Novoa A, Codourey J, Nguyen Huynh TH, Gonzalez F, Milinkovitch MC, Mallo M, Duboule D. 2016. Reorganisation of Hoxd regulatory landscapes during the evolution of a snake-like body plan ed. R. Krumlauf. *eLife* **5**: e16087. https://doi.org/10.7554/eLife.16087 (Accessed April 16, 2020).

Guo T, Mandai K, Condie BG, Wickramasinghe SR, Capecchi MR, Ginty DD. 2011. An evolving NGF-Hoxd1 signaling pathway mediates development of divergent neural circuits in vertebrates. *Nat Neurosci* **14**: 31–36. http://www.nature.com/neuro/journal/v14/n1/full/nn.2710.html (Accessed May 26, 2016).

Harmston N, Ing-Simmons E, Tan G, Perry M, Merkenschlager M, Lenhard B. 2017. Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nat Commun* **8**. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5585340/ (Accessed October 8, 2017).

He S, Viso F del, Chen C-Y, Ikmi A, Kroesen AE, Gibson MC. 2018. An axial Hox code controls tissue segmentation and body patterning in Nematostella vectensis. *Science* **361**: 1377–1380. https://science.sciencemag.org/content/361/6409/1377 (Accessed June 15, 2021).

Hérault Y, Beckers J, Kondo T, Fraudeau N, Duboule D. 1998. Genetic analysis of a Hoxd-12 regulatory element reveals global versus local modes of controls in the HoxD complex. *Dev Camb Engl* **125**: 1669–1677.

Hnisz D, Weintraub AS, Day DS, Valton A-L, Bak RO, Li CH, Goldmann J, Lajoie BR, Fan ZP, Sigova AA, et al. 2016. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**: 1454–1458. https://science.sciencemag.org/content/351/6280/1454 (Accessed June 19, 2021).

Holland LZ, Albalat R, Azumi K, Benito-Gutiérrez È, Blow MJ, Bronner-Fraser M, Brunet F, Butts T, Candiani S, Dishaw LJ, et al. 2008. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res* **18**: 1100–1111. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2493399/ (Accessed May 20, 2016).

Holland PWH. 2013. Evolution of homeobox genes. *Wiley Interdiscip Rev Dev Biol* **2**: 31–45. http://onlinelibrary.wiley.com/doi/10.1002/wdev.78/abstract (Accessed March 26, 2016).

Hrycaj SM, Wellik DM. 2016. Hox genes and evolution. *F1000Research* **5**. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4863668/ (Accessed May 28, 2016).

Ibrahim DM, Mundlos S. 2020. Three-dimensional chromatin in disease: What holds us together and what drives us apart? *Curr Opin Cell Biol* **64**: 1–9. https://www.sciencedirect.com/science/article/pii/S0955067420300041 (Accessed June 19, 2021).

Izpisúa-Belmonte JC, Ede DA, Tickle C, Duboule D. 1992. The mis-expression of posterior Hox-4 genes in talpid (ta3) mutant wings correlates with the absence of anteroposterior polarity. *Dev Camb Engl* **114**: 959–963.

Jerkovic I, Cavalli G. 2021. Understanding 3D genome organization by multidisciplinary methods. *Nat Rev Mol Cell Biol*.

Kasahara M. 2007. The 2R hypothesis: an update. *Curr Opin Immunol* **19**: 547–552. http://www.sciencedirect.com/science/article/pii/S0952791507001239 (Accessed October 2, 2020).

Kentepozidou E, Aitken SJ, Feig C, Stefflova K, Ibarra-Soria X, Odom DT, Roller M, Flicek P. 2020. Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome Biol* **21**: 5. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1894-x (Accessed February 3, 2021).

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* **188**: 107–116. https://science.sciencemag.org/content/188/4184/107 (Accessed June 2, 2021).

Kondo T, Zákány J, Duboule D. 1998. Control of colinearity in AbdB genes of the mouse HoxD complex. *Mol Cell* **1**: 289–300.

Kragesteen BK, Spielmann M, Paliou C, Heinrich V, Schöpflin R, Esposito A, Annunziatella C, Bianco S, Chiariello AM, Jerković I, et al. 2018. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. *Nat Genet* **50**: 1463–1473. https://www.nature.com/articles/s41588-018-0221-x (Accessed April 20, 2020).

Krumlauf R. 1994. Hox genes in vertebrate development. *Cell* **78**: 191–201. http://www.sciencedirect.com/science/article/pii/0092867494902909 (Accessed March 11, 2015).

Lemons D, McGinnis W. 2006. Genomic Evolution of Hox Gene Clusters. *Science* **313**: 1918–1922. http://www.sciencemag.org/content/313/5795/1918 (Accessed May 29, 2015).

Levine M, Rubin GM, Tjian R. 1984. Human DNA sequences homologous to a protein coding region conserved between homeotic genes of Drosophila. *Cell* **38**: 667–673.

Levine M, Tjian R. 2003. Transcription regulation and animal diversity. *Nature* **424**: 147–151. https://www.nature.com/articles/nature01763 (Accessed September 16, 2020).

Lewis EB. 1978. A gene complex controlling segmentation in Drosophila. *Nature* **276**: 565–570.

Li Z, Tiley GP, Galuska SR, Reardon CR, Kidder TI, Rundell RJ, Barker MS. 2018. Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proc Natl Acad Sci* **115**: 4713–4718.

Lonfat N, Montavon T, Darbellay F, Gitto S, Duboule D. 2014. Convergent evolution of complex regulatory landscapes and pleiotropy at Hox loci. *Science* **346**: 1004–1006. http://science.sciencemag.org/content/346/6212/1004 (Accessed May 24, 2016).

Long HK, Prescott SL, Wysocka J. 2016. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**: 1170–1187. http://www.sciencedirect.com/science/article/pii/S009286741631251X (Accessed April 20, 2020).

Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, et al. 2015. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161**: 1012–1025. http://www.sciencedirect.com/science/article/pii/S0092867415003773 (Accessed April 20, 2020).

Maeso I, Tena JJ. 2016. Favorable genomic environments for cis-regulatory evolution: A novel theoretical framework. *Semin Cell Dev Biol* **57**: 2–10. http://www.sciencedirect.com/science/article/pii/S1084952115300161 (Accessed April 20, 2020).

Mallo M, Wellik DM, Deschamps J. 2010. Hox Genes and Regional Patterning of the Vertebrate Body Plan. *Dev Biol* **344**: 7–15. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2909379/ (Accessed May 8, 2016).

Marlétaz F, Firbas PN, Maeso I, Tena JJ, Bogdanovic O, Perry M, Wyatt CDR, Calle-Mustienes E de la, Bertrand S, Burguera D, et al. 2018. Amphioxus functional

genomics and the origins of vertebrate gene regulation. *Nature* **564**: 64. https://www.nature.com/articles/s41586-018-0734-6 (Accessed January 16, 2019).

Martindale MQ. 2004. Investigating the origins of triploblasty: `mesodermal' gene expression in a diploblastic animal, the sea anemone Nematostella vectensis (phylum, Cnidaria; class, Anthozoa). *Development* **131**: 2463–2474. http://dev.biologists.org/cgi/doi/10.1242/dev.01119 (Accessed March 18, 2015).

Martinez-Morales JR. 2016. Toward understanding the evolution of vertebrate gene regulatory networks: comparative genomics and epigenomic approaches. *Brief Funct Genomics* **15**: 315–321. https://doi.org/10.1093/bfgp/elv032 (Accessed February 10, 2021).

McGinnis W, Garber RL, Wirz J, Kuroiwa A, Gehring WJ. 1984. A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. *Cell* **37**: 403–408.

Meyerowitz EM. 1999. Plants, animals and the logic of development. *Trends Cell Biol* **9**: M65–M68. https://www.cell.com/trends/cell-biology/abstract/S0962-8924(99)01649-9 (Accessed August 13, 2021).

Mir M, Bickmore W, Furlong EEM, Narlikar G. 2019. Chromatin topology, condensates and gene regulation: shifting paradigms or just a phase? *Development* **146**. https://dev.biologists.org/content/146/19/dev182766 (Accessed April 21, 2020).

Misteli T, Finn EH. 2021. Chromatin architecture is a flexible foundation for gene expression. *Nat Genet* 1–2. https://www.nature.com/articles/s41588-021-00813-2 (Accessed April 6, 2021).

Montavon T, Soshnikova N, Mascrez B, Joye E, Thevenet L, Splinter E, de Laat W, Spitz F, Duboule D. 2011. A Regulatory Archipelago Controls Hox Genes Transcription in Digits. *Cell* **147**: 1132–1145. http://www.sciencedirect.com/science/article/pii/S0092867411012736 (Accessed June 22, 2015).

Nakatani Y, Shingate P, Ravi V, Pillai NE, Prasad A, McLysaght A, Venkatesh B. 2021. Reconstruction of proto-vertebrate, proto-cyclostome and proto-gnathostome genomes provides new insights into early vertebrate evolution. *Nat Commun* **12**: 4489. https://www.nature.com/articles/s41467-021-24573-z (Accessed July 29, 2021).

Noordermeer D, Duboule D. 2013. Chapter Four - Chromatin Architectures and Hox Gene Collinearity. In *Current Topics in Developmental Biology* (ed. E. Heard), Vol. 104 of *Epigenetics and Development*, pp. 113–148, Academic Press

https://www.sciencedirect.com/science/article/pii/B9780124160279000048 (Accessed June 5, 2021).

Nora EP, Goloborodko A, Valton A-L, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG. 2017. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**: 930-944.e22. http://www.sciencedirect.com/science/article/pii/S0092867417305317 (Accessed April 20, 2020).

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, Berkum NL van, Meisig J, Sedat J, et al. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**: 381. https://www.nature.com/articles/nature11049 (Accessed January 22, 2018).

Ohno,S. 1970. Evolution by Gene duplication.

Oliver G, Sidell N, Fiske W, Heinzmann C, Mohandas T, Sparkes RS, De Robertis EM. 1989. Complementary homeo protein gradients in developing limb buds. *Genes Dev* **3**: 641–650. http://www.genesdev.org/cgi/doi/10.1101/gad.3.5.641 (Accessed June 2, 2021).

Oudelaar AM, Higgs DR. 2021. The relationship between genome structure and function. *Nat Rev Genet* **22**: 154–168. https://www.nature.com/articles/s41576-020-00303-x (Accessed June 20, 2021).

Ovcharenko I, Loots GG, Nobrega MA, Hardison RC, Miller W, Stubbs L. 2005. Evolution and functional classification of vertebrate gene deserts. *Genome Res* **15**: 137–145. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC540279/ (Accessed October 2, 2020).

Paliou C, Guckelberger P, Schöpflin R, Heinrich V, Esposito A, Chiariello AM, Bianco S, Annunziatella C, Helmuth J, Haas S, et al. 2019. Preformed chromatin topology assists transcriptional robustness of Shh during limb development. *Proc Natl Acad Sci U S A* **116**: 12390–12399.

Pastrana CC, DeBiasse MB, Ryan JF. 2019. Sponges Lack ParaHox Genes. *Genome Biol Evol* **11**: 1250–1257. https://doi.org/10.1093/gbe/evz052 (Accessed June 15, 2021).

Rao SSP, Huang S-C, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon K-R, Sanborn AL, Johnstone SE, Bascom GD, Bochkov ID, et al. 2017. Cohesin Loss Eliminates All Loop Domains. *Cell* **171**: 305-320.e24. http://www.sciencedirect.com/science/article/pii/S0092867417311200 (Accessed April 20, 2020).

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. 2014. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**: 1665–1680. http://www.cell.com/cell/abstract/S0092-8674(14)01497-4 (Accessed October 15, 2017).

Ravi V, Lam K, Tay B-H, Tay A, Brenner S, Venkatesh B. 2009. Elephant shark (Callorhinchus milii) provides insights into the evolution of Hox gene clusters in gnathostomes. *Proc Natl Acad Sci* **106**: 16327–16332. https://www.pnas.org/content/106/38/16327 (Accessed February 26, 2019).

Reddy PC, Unni MK, Gungi A, Agarwal P, Galande S. 2015. Evolution of Hox-like genes in Cnidaria: Study of Hydra Hox repertoire reveals tailor-made Hox-code for Cnidarians. *Mech Dev* **138**: 87–96. https://www.sciencedirect.com/science/article/pii/S0925477315300228 (Accessed June 15, 2021).

Rentzsch F, Holstein TW. 2018. Making head or tail of cnidarian hox gene function. *Nat Commun* **9**: 2187. https://www.nature.com/articles/s41467-018-04585-y (Accessed June 15, 2021).

Richardson MK, Allen SP, Wright GM, Raynaud A, Hanken J. 1998. Somite number and vertebrate evolution. *Development* **125**: 151–160. http://dev.biologists.org/content/125/2/151 (Accessed May 1, 2016).

Rittmeyer EN, Allison A, Gründler MC, Thompson DK, Austin CC. 2012. Ecological Guild Evolution and the Discovery of the World's Smallest Vertebrate. *PLOS ONE* **7**: e29797. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0029797 (Accessed June 4, 2021).

Rodríguez-Carballo E, Lopez-Delisle L, Willemin A, Beccari L, Gitto S, Mascrez B, Duboule D. 2020. Chromatin topology and the timing of enhancer function at the HoxD locus. *Proc Natl Acad Sci*. https://www.pnas.org/content/early/2020/11/20/2015083117 (Accessed December 8, 2020).

Rodríguez-Carballo E, Lopez-Delisle L, Yakushiji-Kaminatsui N, Ullate-Agote A, Duboule D. 2019. Impact of genome architecture on the functional activation and repression of Hox regulatory landscapes. *BMC Biol* **17**: 55. https://doi.org/10.1186/s12915-019-0677-x (Accessed June 21, 2021).

Rodríguez-Carballo E, Lopez-Delisle L, Zhan Y, Fabre PJ, Beccari L, El-Idrissi I, Huynh THN, Ozadam H, Dekker J, Duboule D. 2017. The HoxD cluster is a dynamic and

resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev* **31**: 2264–2281. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5769770/ (Accessed May 3, 2020).

Roscito JG, Sameith K, Parra G, Langer BE, Petzold A, Moebius C, Bickle M, Rodrigues MT, Hiller M. 2018. Phenotype loss is associated with widespread divergence of the gene regulatory landscape in evolution. *Nat Commun* **9**: 4737. https://www.nature.com/articles/s41467-018-07122-z (Accessed August 28, 2020).

Schep R, Necsulea A, Rodríguez-Carballo E, Guerreiro I, Andrey G, Huynh THN, Marcet V, Zákány J, Duboule D, Beccari L. 2016. Control of Hoxd gene transcription in the mammary bud by hijacking a preexisting regulatory landscape. *Proc Natl Acad Sci* **113**: E7720–E7729. https://www.pnas.org/content/113/48/E7720 (Accessed February 21, 2019).

Schummer M, Scheurlen I, Schaller C, Galliot B. 1992. HOM/HOX homeobox genes are present in hydra (Chlorohydra viridissima) and are differentially expressed during regeneration. *EMBO J* **11**: 1815–1823. https://www.embopress.org/doi/abs/10.1002/j.1460-2075.1992.tb05233.x (Accessed June 15, 2021).

Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, Fonseca NA, Huber W, H. Haering C, Mirny L, et al. 2017. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**: 51–56. https://www.nature.com/articles/nature24281 (Accessed April 1, 2019).

Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, et al. 2012. A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**: 116–120. http://www.nature.com/articles/nature11243 (Accessed April 8, 2019).

Shu D-G, Morris SC, Han J, Zhang Z-F, Yasui K, Janvier P, Chen L, Zhang X-L, Liu J-N, Li Y, et al. 2003. Head and backbone of the Early Cambrian vertebrate Haikouichthys. *Nature* **421**: 526–529. https://www.nature.com/articles/nature01264 (Accessed August 9, 2021).

Simakov O, Marlétaz F, Yue J-X, O'Connell B, Jenkins J, Brandt A, Calef R, Tung C-H, Huang T-K, Schmutz J, et al. 2020. Deeply conserved synteny resolves early events in vertebrate evolution. *Nat Ecol Evol* 1–11. https://www.nature.com/articles/s41559-020-1156-z (Accessed April 23, 2020).

Smith JJ, Kuraku S, Holt C, Sauka-Spengler T, Jiang N, Campbell MS, Yandell MD, Manousaki T, Meyer A, Bloom OE, et al. 2013. Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. *Nat Genet* **45**: 415–421. https://www.nature.com/articles/ng.2568 (Accessed August 3, 2021).

Soshnikova N, Dewaele R, Janvier P, Krumlauf R, Duboule D. 2013. Duplications of hox gene clusters and the emergence of vertebrates. *Dev Biol* **378**: 194–199. https://www.sciencedirect.com/science/article/pii/S0012160613001334 (Accessed June 5, 2021).

Spitz F, Furlong EEM. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**: 613–626. https://www.nature.com/articles/nrg3207 (Accessed April 20, 2020).

Spitz F, Gonzalez F, Duboule D. 2003. A Global Control Region Defines a Chromosomal Regulatory Landscape Containing the HoxD Cluster. *Cell* **113**: 405–417. http://www.sciencedirect.com/science/article/pii/S0092867403003106 (Accessed May 6, 2016).

Spitz F, Gonzalez F, Peichel C, Vogt TF, Duboule D, Zákány J. 2001. Large scale transgenic and cluster deletion analysis of the HoxD complex separate an ancestral regulatory module from evolutionary innovations. *Genes Dev* **15**: 2209–2214. http://genesdev.cshlp.org/content/15/17/2209 (Accessed April 5, 2015).

Spitz F, Herkenne C, Morris MA, Duboule D. 2005. Inversion-induced disruption of the Hoxd cluster leads to the partition of regulatory landscapes. *Nat Genet* **37**: 889–893. https://www.nature.com/articles/ng1597 (Accessed June 17, 2021).

Symmons O, Uslu VV, Tsujimura T, Ruf S, Nassari S, Schwarzer W, Ettwiller L, Spitz F. 2014. Functional and topological characteristics of mammalian regulatory domains. *Genome Res* **24**: 390–400. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3941104/.

Touceda-Suárez M, Kita EM, Acemel RD, Firbas PN, Magri MS, Naranjo S, Tena JJ, Gómez-Skarmeta JL, Maeso I, Irimia M. 2020. Ancient Genomic Regulatory Blocks Are a Source for Regulatory Gene Deserts in Vertebrates after Whole-Genome Duplications. *Mol Biol Evol* **37**: 2857–2864. https://doi.org/10.1093/molbev/msaa123 (Accessed February 2, 2021).

Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, Hadjur S. 2015. Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. *Cell Rep* **10**: 1297–1309.

http://www.sciencedirect.com/science/article/pii/S2211124715001126 (Accessed March 27, 2019).

Woltering JM, Noordermeer D, Leleu M, Duboule D. 2014. Conservation and Divergence of Regulatory Strategies at Hox Loci and the Origin of Tetrapod Digits. *PLoS Biol* **12**. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3897358/ (Accessed February 19, 2015).

Wood TWP, Nakamura T. 2018. Problems in Fish-to-Tetrapod Transition: Genetic Expeditions Into Old Specimens. *Front Cell Dev Biol* **0**. https://www.frontiersin.org/articles/10.3389/fcell.2018.00070/full (Accessed August 9, 2021).

York JR, Yuan T, McCauley DW. 2020. Evolutionary and Developmental Associations of Neural Crest and Placodes in the Vertebrate Head: Insights From Jawless Vertebrates. *Front Physiol* **11**: 986. https://www.frontiersin.org/article/10.3389/fphys.2020.00986 (Accessed August 14, 2021).

Zákány J, Duboule D. 1999. Hox genes and the making of sphincters. *Nature* **401**: 761–762. http://www.nature.com/nature/journal/v401/n6755/full/401761a0.html (Accessed May 29, 2016).

Zákány J, Gérard M, Favier B, Duboule D. 1997. Deletion of a HoxD enhancer induces transcriptional heterochrony leading to transposition of the sacrum. *EMBO J* **16**: 4393–4402. http://onlinelibrary.wiley.com/doi/10.1093/emboj/16.14.4393/abstract (Accessed April 3, 2016).