Trends in Plant Science

Cell Press REVIEWS

Review

Copy Number Variation in Domestication

Zoe N. Lye[1] and Michael D. Purugganan[1,2,*]

Domesticated plants have long served as excellent models for studying evolution. Many genes and mutations underlying important domestication traits have been identified, and most causal mutations appear to be SNPs. Copy number variation (CNV) is an important source of genetic variation that has been largely neglected in studies of domestication. Ongoing work demonstrates the importance of CNVs as a source of genetic variation during domestication, and during the diversification of domesticated taxa. Here, we review how CNVs contribute to evolutionary processes underlying domestication, and review examples of domestication traits caused by CNVs. We draw from examples in plant species, but also highlight cases in animal systems that could illuminate the roles of CNVs in the domestication process.

Highlights

Whole-genome resequencing, pangenomics, and developing computational methods have allowed characterization of CNVs in diverse species.

Loss-of-function CNVs can cause some of the critical domestication traits in plants, whereas other CNVs are associated with postdomestication diversification traits, such as environmental adaptation, disease resistance, fruit size, and cultural preferences.

An exhaustive table of characterized CNVs associated with domestication phenotypes in a plant and animal systems is included.

Domestication is a Coevolutionary Process

Domestication is an evolutionary process that arises from coevolutionary interactions where one species controls the reproduction and dispersal of another species for the benefit of the former. Human-associated domestication as an evolutionary process began in the Paleolithic and continued into the Neolithic, with the shift of hunter-gatherers to pastoralists and farmers beginning ~12 000 years ago, leading to the evolution of hundreds of crop plant species [1]. Moreover, domestication also occurred in animals, and there are dozens of known domesticated livestock and pet species [2]. It is now generally thought that domestication was a protracted process that unfolded over thousands of years [3,4] and, it was during this period, that genetic changes led to adaptation to agricultural environments and differentiation from wild ancestors.

The early evolution of domesticated species occurs in two distinct phases: (i) initial domestication, where control over reproduction and dispersal is established, resulting in the origin of the new domesticated species; and (ii) diversification and/or improvement, where the domesticated species develops local or population-specific adaptations to different environments or cultural preferences as it spreads from its center of origin [3–5]. Many of the adaptive traits arising during this process may have evolved under the process termed ‘unconscious selection’, which acts similar to natural selection because incipient domesticates adapt to living in human-associated environments [1,2]. Nevertheless, many key traits, particularly those associated with diversification, may have evolved under more intense selection.

Most studies on the evolutionary genetics of domestication have used SNPs to examine population relationships and to identify causal genetic variants often through genetic mapping and genome-wide association studies. The role of CNVs in the evolution of domesticated species is not as well appreciated. In recent years, as whole-genome sequencing methods have allowed the genome-wide characterization of CNVs, they have become the subject of increased interest, broadening our understanding of the genetic basis of evolution. Here, we review the role of CNVs in domestication, focusing primarily on plant species, but also providing

[1]Center for Genomics and Systems Biology, 12 Waverly Place, New York University, New York, NY 10003, USA
[2]Center for Genomics and Systems Biology, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, United Arab Emirates

*Correspondence: mp132@nyu.edu (M.D. Purugganan).

 https://doi.org/10.1016/j.tplants.2019.01.003
© 2019 Elsevier Ltd. All rights reserved.

CellPress REVIEWS

examples from domesticated animal species that could point to contrasting patterns between these two groups.

Copy Number Variation

CNVs are polymorphisms within species in which sections of a genome differ in copy number between individuals, and include deletions, duplications, or **amplifications** (see Glossary) of DNA sequence. Originally, CNVs were only thought of in terms of copy changes in functional genetic features. Today, many researchers adopt a more expansive definition in light of the ability to discover gains and losses of genomic material in an unbiased genome-wide manner (Box 1). This can the include transposable elements and noncoding sequences. The definition of a CNV continues to be somewhat arbitrary and is often conflated with the terms 'segmental duplication' or 'structural variant'. The defined minimum length of a CNV is typically 1 kb, although many studies include smaller variants of as few as 50 base pairs (bps) [6,7]. Nevertheless, CNVs that include functional sequences continue to be of most interest to researchers.

Box 1. CNV Detection Methods

For reviews on major methods for CNV detection applied to domesticated species, see [7]. For reviews of CNV detection from next-generation sequencing data, see [103,104].

Array comparative genome hybridization (aCGH) is based on the comparison of fluorescence signals of a test and reference sample hybridized to a microarray of tiled probes covering an entire genome. The use of smaller probes increases the specificity of CNV detection in this method; however, aCGH (is more accurate in detecting deletions than duplications [7,105]. SNP microarrays are also applied to CNV detection by comparing probe intensities across samples. They are also able to distinguish CNV alleles because they can use allele-specific probes [7].

Next-generation sequencing (NGS)-based methods fall into three major categories: read-depth (RD), read-pair (RP), and split-read (SR) methods [7,103]. RD methods detect CNVs by comparing normalized read depth from short-read sequence data aligned to a reference genome. Low or zero RD is interpreted as a deletion and increased RD is interpreted as an increase in copy number. RP methods are based on the idea that read pairs should map to a reference separated by approximately the same distances as the insert size. If read-pairs map farther away from each other than expected, a deletion is detected; if they are too close together, an insertion is detected. SR methods use paired-end reads and detect CNVs by aberrant mapping to a reference genome. For example, when only half of a read-pair maps to a genome, a CNV breakpoint is identified. Whole-exome sequencing data are also applied to CNV discovery using a RD approach to identify CNVs [104]. Additionally, local realignments are also used to refine the identification of CNV breakpoints and infer structure of CNVs [7].

Each method comes with a different set of biases. RP methods are less effective in repetitive regions and their accuracy is dependent on the size of the insert [103,104]. SR methods are biased to detect smaller CNVs [103]. RD methods typically have higher false positive rates and are biased towards detecting large variants [7]. The effectiveness of these methods is also dependent on sample read depth. Due to these shortcomings, CNV studies using NGS data typically combine multiple computational approaches to minimize false positives [7].

There does not appear to be clear methodological standards in the field of CNV discovery. Most, if not all, CNV discovery methods were developed for use in humans, and can be benchmarked against gold standard sets of known human variants. In domesticated species, gold standard CNV sets do not exist to evaluate the efficacy of different methodologies. Rather, researchers rely on simulations to benchmark methods or simply take existing methods at face value. There are more than 50 published methods for detecting CNVs from NGS data. Selecting an appropriate method for a given data set and species is a challenge to anyone designing a CNV study. As multiple high-quality reference genomes are created for domesticated species and third-generation long-read sequencing becomes available, we expect to see an increase in CNV studies and the development of more novel methodologies. Long read sequences have already been used to resolve CNV in tandem repeats where traditional methods are limited [106]. It is critical that new methodologies developed are accurately compared to existing methods to ensure that research is comparable across platforms.

Glossary

Amplification: the same sequence of DNA is duplicated multiple times, typically in tandem.

Chimeric gene: a gene comprising coding sequences derived from two or more other genes.

Experimental evolution: the use of laboratory or controlled field experiments to investigate the processes of evolution. Typically, organisms with short generations times are used to simulate processes that would take longer in larger organisms.

Fixation: increase in frequency of a genetic variant, eventually resulting in all members of a population sharing the same variant at a locus.

Fixation index (F_{ST}): a measure of genetic differentiation between two populations.

Microhomology: identical short DNA sequences, 1–4 bp in length.

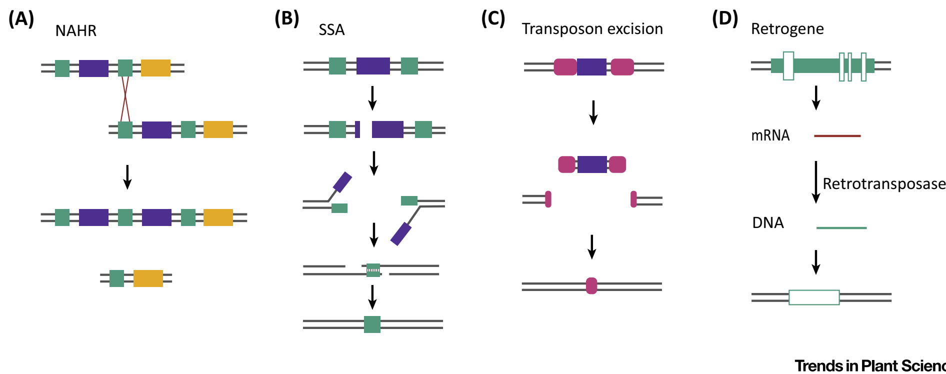
Pan-genome: the entire gene set contained within a species, taking into account PAV between individuals in a species. Not all individuals carry all of the genes in the pan-genome.

Photoperiod: day length; many plants use day length as a signal to enter various stages of the life cycle.

Purifying selection: selection against disadvantageous alleles.

Tandem array: cluster of genes created by repeated duplications.

Vernalization: induction of flowering by prolonged exposure to cold (i.e., winter).



Trends in Plant Science

Figure 1. Mechanisms of Copy Number Variation (CNV) Formation. (A) Nonallelic homologous recombination (NAHR; unequal crossing over): during a recombination-based double-strand break (DSB) repair, a direct repeat, represented in green, is used as homology and incorrectly pairs during crossing over, this causes a reciprocal deletion and duplication of sequence between the repeats (purple). In this scenario, the resulting CNV break point is flanked by tracts of homologous sequence. (B) Single-strand annealing (SSA). During double-strand break repair, the 5′ stands are resected to expose complimentary sequences either side of the break (green). Although this is similar to the micro-homology-mediated end joining repair pathway, SSA requires longer tracts of homology, typically >30 base pairs (bp). This can result in significant deletions of intervening sequence (purple). (D) Transposon excision. Transposons (pink ovals) flank a unique sequence (purple). Both transposons excise simultaneously, removing the unique sequence with them, and can result in a deletion. (D) Retro-gene formation. Retrotransposon activity causes insertion of a coding sequence into the genome (gene is shown in green with white boxes representing introns). mRNA (red) from the gene is reverse transcribed to DNA. This DNA can be occasionally inserted into the genome and become a retrogene, a copy of the original gene lacking introns (green box). These genes can be inserted into another gene, creating a chimeric gene, or become under control of different promoter sequences and take on a new expression regime.

CNVs are formed through a variety of genetic mechanisms (reviewed in [6]). A key mechanism is nonallelic homologous recombination (NAHR) or unequal crossing over, which results from aberrant homology recognition during homology-based DNA repair or meiosis [6] (Figure 1A). CNVs formed by this mechanism are characterized by tracts of homology on either side of the CNV. NAHR is common in repetitive regions and an important source of tandem duplications and deletions [8]. Another mechanism is single-strand annealing (SSA), which is a double-strand break repair process where broken ends are joined by annealing at homologies >30 bp in length, which can result in significant deletions [6] (Figure 1B).

Transposable elements are also a source of CNVs; they can result in copy number change by capturing DNA segments during excision and moving or deleting DNA segments [6] (Figure 1C). Retrotransposon activity can also create CNVs through retrogenes; these are DNA insertions into the genome resulting from reverse-transcribed mRNA that might take on a new function or form a **chimeric gene** [9] (Figure 1D). For example, the *sun* locus in tomato is a retrotransposon-mediated gene duplication that places the *SUN* gene under a different regulatory element, altering fruit development to result in an oval fruit [10].

CNVs can also arise following polyploidization, when a genome doubles all genes are duplicated; subsequent deletions in either of the subgenomes lead to a change in copy number. Fractionalization of the maize genome has contributed to high intraspecific variation in copy number and presence–absence variation (PAV) [11,12].

Other proposed mechanisms include **microhomology**-mediated break-induced replication (MMBIR), whereby, during meiosis, a replication fork stalls and the lagging strand anneals to a different replication in the vicinity as a result of microhomology, which can lead to complex

rearrangements and duplications [13]. Undoubtedly, there are other stochastic and poorly understood causes of CNV formation, especially ones that involve larger CNVs (for a comprehensive review of these mechanisms, see [6]).

CNVs Are Generally Deleterious . . .

CNVs have become the subject of increased interest, broadening our understanding of the genetic basis of evolution. CNVs are thought to be generally deleterious and subject to **purifying selection** and, thus, affect coding sequences less frequently than noncoding sequences [14–16]. Deletion CNVs can lead to loss of function (LOF), whereas duplication CNVs affecting entire protein-coding genes can be deleterious if they affect dosage-sensitive genes [14,17]. Simulation of the effects of genic CNVs in regulatory networks demonstrated that increases in gene copy number by one or two copies can have large effects on overall expression patterns due to regulatory feedbacks [18]. CNVs have been identified as expression quantitative trait loci (eQTLs), further demonstrating their role in altering gene expression [19,20].

Given the generally deleterious effects of CNVs, it would be expected that CNVs that do affect gene expression should be restricted to functional classes that can tolerate expression changes without costs to fitness. Thus, most genic CNVs are predicted to occur in lowly expressed genes at the periphery of gene regulatory and gene interaction networks, where change in copy number is less impactful. Dopman and Hartl measured this in *Drosophila* and found significantly lower representation of deletion CNV genes among genes with known interactions [21]. They also measured the ratio between nonsynonymous (Ka) and synonymous site (Ks) mutations in open reading frames of CNV genes and found that CNV genes had a higher ratio than did non-CNV genes, suggesting that CNV genes are under relaxed selective constraint [21]. Keel *et al*. extended the investigation of CNVs in interaction networks to cows by quantifying the number of interactions of CNV genes in a protein–protein interaction network [22]. They demonstrated that CNV genes were likely to have fewer network connections than were non-CNV genes, supporting the prediction that CNV genes are functionally constrained and tend to occur at the periphery of interaction networks.

CNVs May Have a Role in Rapid Adaptation under Strong Selective Pressure

While CNVs are generally deleterious, they also appear to be a key mechanism that can enable adaptation during a period of strong selection. This phenomenon is observed in **experimental evolution** of microbes under nutrient limitation, where spontaneous duplication of nutrient transporters repeatably occurs, conferring adaptation to the nutrient-limited environment. In yeast under glucose-limited conditions, for example, amplifications of the *HXT6* and *HXT7* genes, encoding high-affinity glucose transporters, were observed; under sulfate limitation, there was amplification of *SUL1*, which encodes a high-affinity sulfate transporter [23].

This effect has also been observed in multicellular systems. An experimental evolution study of *Arabidopsis thaliana* grown under stress conditions of high heat and salicylic acid showed increased CNV formation [24]. Similar results were found in an experimental evolution study in *Caenorhabditis elegans* that selected for recovered fecundity following inbreeding and mutagen application, where an increase in the frequency of copy number change was observed during the adaptive recovery stage [25]. Interestingly, CNVs in replicate populations were identified at the same genome regions but had different breakpoints, suggesting recurrent adaptation [25].

CellPress REVIEWS

In the natural environment, adaptive CNVs are also found in response to strong selective pressures. For example, amplifications of P540 genes have conferred insecticide resistance in aphids and multiple disease vector mosquito species [26–28]. Given that many species undergo a period of strong selective pressure during domestication, typically in the postdomestication diversification stage, CNVs could have been an important source of genetic diversity underlying adaptation during domestication.

CNVs Are Widespread in Domesticated Species

The results of experimental evolution experiments suggest that CNVs contribute to the rapid adaptation associated with domestication and during population expansion of the domesticated species. Advances in detection methodologies (summarized in Box 1), reduced sequencing costs, and proliferation of sequencing data have expanded CNV studies, and CNVs have been described in most major crop plant species, including rice, maize, potato, soybean, barley, cucumber, melon, apple, and grapevine (Table 1, Key Table) [29–37]. Not surprisingly, they have also been examined in domesticated animal species, such as silkworm, sheep, goat, pig, chicken, cow, horse, and dog (Table 1) [22,38–43]. These studies demonstrate that CNVs are a pervasive source of genetic variation in domesticated taxa, and examples from both plants and animals serve to highlight both common and contrasting features of CNVs in both groups.

Early studies of CNVs in domesticated species used few samples, although they nevertheless provided key insights. In rice (*Oryza sativa*), for example, whole-genome comparisons of two cultivars found 641 CNVs ranging in size from 1.1 kb to 180.7 kb [44]. An analysis of two inbred lines in maize (*Zea mays)* found ~400 genomic regions exhibiting duplications and pervasive PAV affecting more than 700 genes [45].

However, increasing sample sizes resulted in more thorough catalogs of CNVs and other structural variants. Later analysis of 11 maize and 14 wild relative teosinte individuals, for example, found 3889 CNVs, most of which were segregating in both maize and teosinte [46]. In the case of rice, a recent study of 3010 rice varieties identified thousands of deletions and hundreds of duplications affecting between 100 bp and 1 Mb [47]. Indeed, there are often significant inconsistencies in the results of CNV analyses from different studies within the same species due to differences in sample sizes, breeds used, and methodologies of CNV detection (Box 1) [22,48,49]. This was highlighted in a recent analysis, albeit in a domesticated animal species, which characterized CNVs in European cattle (*Bos taurus*) populations and compared the results to 18 previous studies [39]. Prior studies had identified between 27 and 3438 CNVRs and, of those data sets, 6–46% of CNVRs overlapped with CNVRs discovered in the present study [39]. Altogether, this analysis indicated that CNVs may affect as much as 63 Mb of the genome, and that cattle have a higher level of CNV diversity than a single study would predict [39].

The role that CNVs have in the evolution of domesticated taxa is becoming clear. Over past three decades, considerable effort has been applied to identify the causal mutations and genes associated with domestication and diversification traits [1,3]. We compiled an exhaustive list of genes from the literature and found 39 examples where CNVs appear to have a role in trait evolution in plant and animal domesticated species (Table 1).

The size of these CNVs associated with domestication and diversification ranged from ~1 kb to ~1 Mb. Plant domestication CNVs affected both domestication and diversification traits, whereas animal CNVs were all associated with postdomestication diversification traits. Plant

Key Table

Table 1. Examples of CNVs Affecting Domestication Traits[a]

Species	Type	Locus	Phenotype	Description	Trait type	Refs
Wheat	mCNV: ~25 kb	*Ppd-B1*	early flowering	Pseudo-response regulator (*Ppd-B1*)	Diversification	[62]
	Deletion: 2 kb	*Ppd-D1a*	Photoperiod insensitivity (short day growth)	Pseudo-response regulator (*Ppd-D1a*)	Diversification	[79]
	mCNV: ~30 kb	*Vrn-A1*	Increased vernalization requirement	MADS-box transcription factor	Diversification	[62]
	mCNV[b]	Fr-A2	Frost resistance	Transcription factor, C-repeat Binding Factor (*CBF-A14*)	Diversification	[80]
	Duplication: >1 mb	*Rht-D1c*	Dwarf phenotype; increased yield	DELLA protein, gibberellic acid insensitive	Diversification	[81]
Rice	Insertion[b]	*Sub1a*	Submergence tolerance	Ethylene receptor	Diversification	[74]
	Duplication: 17.1 kb	*GL7*	Grain length	Uncharacterized gene function, homologous to LONGIFOLIA in *Arabidopsis*	Domestication/ Diversification	[68]
	Deletion: 2 alleles: 950 and 1212 bp	*qSW5/ GSE5*	Grain width	*GSE5*, plasma membrane-associated protein	Domestication/ Diversification	[82]
	Insertion: 20.9 kb	*q-AG-9-2*	Anaerobic germination tolerance	Trehalose-6-phosphate phosphatase (*OsTPP7*), sugar signaling and metabolism	Diversification	[83]
	Insertion[b]	*SNORKEL1* *SNORKEL2*	Submergence tolerance	Ethylene response factor; transcription factor	Diversification	[84]
	Insertion: 90 kb	*Pup1*	Low phosphorous tolerance	*PSTOL* receptor-like cytoplasmic- kinase	Diversification	[85]
African Rice	Deletion: 30 kb	*sh1*	Shattering	*YABBY* transcription factor	Domestication	[86]
Sorghum	Deletion: 2.2 kb	*Sh1*	*Seed shattering*	*YABBY*-like transcription factor	Domestication	[52]
Soybean	mCNV: 31 kb	*Rhg1-b*	Resistance to cyst nematode disease	Multiple genes: alpha-SNAP involved in snare membrane traffic, wound-inducible protein 12 (*WI12*), a predicted amino acid transporter	Diversification	[87]
Maize	mCNV[b]	*Rp1*	Resistance to leaf rust disease	Cluster of leucine-rich repeat high CN haplotypes	Diversification	[88]
	Insertion: 147 kb	*ZmWAK*	Resistance to head smut disease	Multiple receptor-like kinase alleles	Diversification	[89]
	mCNV: 30 kb	*MATE1*	Aluminum toxicity resistance	Multidrug and toxic compound extrusion 1 (*MATE1*)	Diversification	[67]
	Duplication: ~1.5 kb	*Tunicate (TU)*	Pod corn	*ZMM19*MADS-box gene	Diversification	[69]
Barley	mCNV[b]	*Bot1*	Boron toxicity resistance	Boron efflux transporter (*Bot1*)	Diversification	[66]
	mCNV[b]: ~6 kb	*HvFT1*	Flowering time	Mobile florigen signaling protein	Diversification	[90]
	Duplication: 22 kb	*VRN-H1*	Freezing tolerance	C-repeat binding factors (*CBF2A-CBF4B*)	Diversification	[91]

CellPress REVIEWS

Table 1. (continued)

Species	Type	Locus	Phenotype	Description	Trait type	Refs
Cucumber	Duplication: 30.2 kb	*Female* (F) locus	Gynoecy	Multiple genes, all likely flowering regulatory: *aminocyclopropane-1-carboxylic acid synthase gene* (*ACS1*); ethylene synthesis; truncated myb transcription factor (Csa6G496960); branched-chain amino acid aminotransferase (Csa6G496970)	Diversification	[29]
Tomato	Retrogene insertion: 24.7 kb	*SUN*	Elongated fruit shape	IQ67 domain-containing family, function uncharacterized	Diversification	[10]
	Deletion: 14 kb	*CSR-D*	Fruit weight	Truncated cell size regulator (*CSR-D*), uncharacterized protein	Domestication/ Diversification	[92]
Common bean	Deletion allele: 5840 bp Insertion allele: 4171 bp	*PvTFL1y*	Determinate growth	Transcription factor controls switch from vegetative to flowering state	Domestication	[53]
Sheep	Duplication > 100 kb	Ovine *ASIP, AHCY*	White coat	Agouti signaling protein (*ASIP*), *S*-adenosylhomocysteine (*AHCY*), and itchy homolog E3 ubiquitin protein ligase promoter (ITCH)	Diversification	[54]
Goat	Duplication 190 kbOvine kb	*ASIP, AHCY*	White coat	*ASIP*, *AHCY*	Diversification	[55]
Pig	Duplication: 450 kb	*KIT*	Coat Color	Mast/stem cell growth factor receptor linked to tyrosine kinase receptor genes	Diversification	[93]
Cattle	Duplication: ~480 kb	*KIT*	White coat	Mast/stem cell growth factor receptor	Diversification	[94]
	mCNV: ~1562 bp	*CYP4A11*	Body fat	Major lauric acid (medium-chain fatty acid) omega hydroxylase, lipogenesis	Diversification	[95]
Dog	Duplication: ~133 kb	*FGF3, FGF4, FGF19, ORAOV1*	Ridgeback	*FGF*, embryonic development; oral cancer overexpressed (ORAOV1;) function uncharacterized	Diversification	[56]
	Retrogene insertion: 5 kb	*fgf4*	Chondrodysplasia (short legs)	*FGF4* retrogene	Diversification	[96]
	mCNV[b]	*AMYB2*	Starch diet	*AMYB2*, pancreatic amylase	Diversification	[77]
	Duplication: 98 kb	Intergenic region	Blue eyes	Intergenic region adjacent to Hox gene ALX4, which has role in eye development	Diversification	[97]
Chicken	Amplification: 3.2 kb	*SOX5*	Peacomb (cold tolerance)	Intron 1 of *SOX5* is a SRY-related HMG box family of transcription factors	Diversification	[98]
	Duplication: 176 kb	*PRLR* and *SPEF2*	Late feathering	Prolactin receptor (*PRLR*); inhibits follicle activation; sperm flagellar protein 2 (*SPEF2*), thought to be involved in signal transmission	Diversification	[99]
	Duplication: 130 kb	*EDN3*	Fibromelanosis (pigmentation)	Endothelin 3 gene (*EDN3*), receptor, melanoblast/melanocyte mitogen	Diversification	[100]
	Duplication: 20 kb	*Duplex-comb*	Comb shape	Duplication upstream of eomesodermin (*EOMES*), a t-box transcription factor	Diversification	[101]
Silkworm	mCNV[b]	*CBP*	Cocoon color	Carotenoid-binding protein (*CBP*); cystolic transporter of carotenoid pigments	Diversification	[41]

[a]Insertion and deletion alleles are distinguished by the state that is associated with the phenotype.
[b]Size of CNV varies, is not precisely known, or amplification units are of variable size.

CellPress REVIEWS

CNVs found in 14 genes were associated with duplications or amplifications, while 11 genes had CNV insertions or deletions; by contrast, of the 14 animal genes identified, all but one were sequence duplications and/or amplifications. Of the duplications and/or amplifications, 22 were tandem duplications. The prevalence of tandem duplications in known crop CNVs may be because previous QTL-mapping techniques made it easier to find tandem rather than dispersed duplications. Plants also tend to have higher genetic redundancy and, thus, may be more robust to deletion and/or PAV mutations; thus, they may be able to better tolerate the deleterious nature of most CNVs.

There were also some differences in the types of domestication and/or diversification genes affected by CNVs in plants versus animals. Crop CNVs had a more diverse array of functions, with CNV mutations found in transcription factors related to **photoperiod** signaling, development, stress tolerance, and resistance genes (R-genes). By contrast, animal CNVs were largely found in genes that encode growth factors and receptors, and genes related to development.

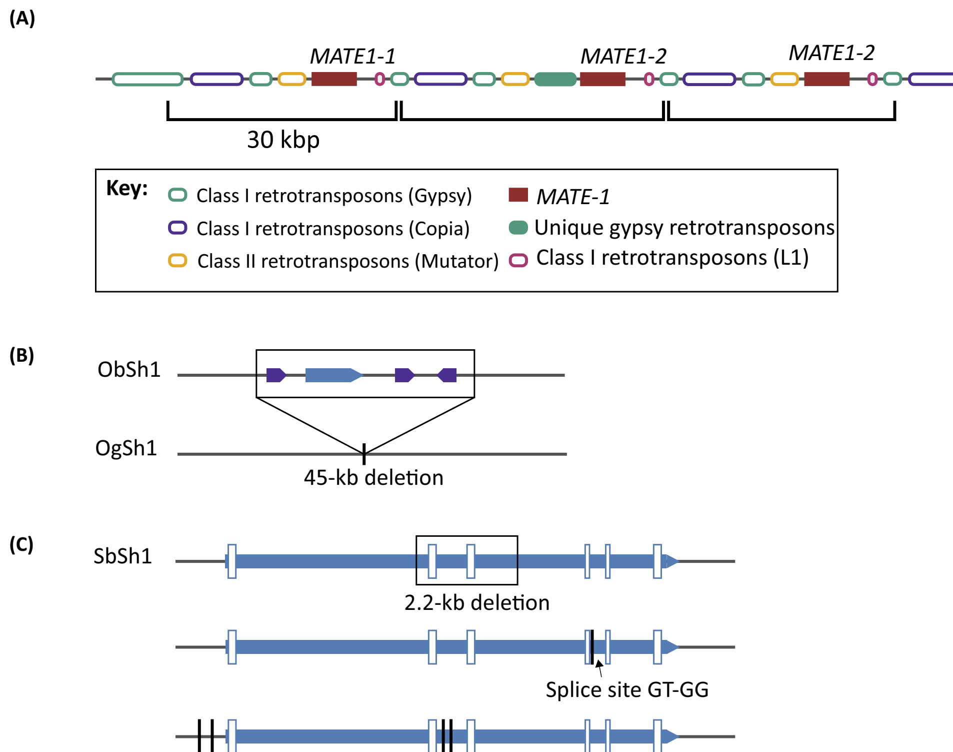
CNVs and Domestication Traits and Genes

Domestication traits are those that distinguish a domesticated species from its wild progenitors, and are the requisite traits for cohabitation with human societies. In crops, common domestication traits include seed nonshattering or suppression of seed dormancy and, in animals, critical changes occur in behavioral traits [1–3]. The genes underlying domestication traits, or 'domestication genes', are thought to have arisen early during the evolution of crop and livestock species (or may even be present in the wild ancestor at low frequencies) [3,4]. Therefore, examining genetic differentiation between wild ancestors and domesticates has long been a strategy to discern the underlying genetic causes of domestication. These approaches are also SNP-centric, and CNVs between wild species and domesticates could further illuminate the genetic basis of domestication.

Domestication traits are common to all members of a domesticated species and, thus, the domestication gene allele is expected to be fixed within a domesticated species; therefore, we do not expect CNVs segregating in a domesticated population to be associated with domestication genes. Nevertheless, causal domestication mutations may have originated from structural variants in an ancestral population. These CNVs would have undergone a transient phase as a CNV on their way to **fixation** during the domestication process. For example, in African rice, *Oryza glaberrima,* LOF in *PROG1*, which governs the transition from prostrate to upright growth, is caused by the deletion of the gene relative to the ancestor [107]. At one time, this deletion must have been a CNV segregating in the population that conferred the adaptive domestication trait, although now can only be considered a fixed deletion mutation relative to the ancestor.

There is evidence for a reduction in CNV diversity in domesticated species compared with wild relatives, which is similar to the reduction in SNP diversity typically observed in domesticates compared with their wild relatives [33,35]. Preferential retention or deletion of genetic sequences from ancestral populations may have been under selection during domestication. A comparison of CNVs between domesticated dogs and wolves, for example, found some CNVs that were strongly differentiated between the two groups [50].

CNVs may be observed in domestication genes within domesticated species if the domestication phenotype is caused by segregating but independent mutations. Loss of seed shattering, for example, is a key domestication trait observed across cereal crops [51], and LOF alleles in the *shattering1* gene (*sh1*) can be caused by both deletions and SNP mutations [52]



Trends in Plant Science

Figure 2. Examples of Copy Number Variations (CNVs). (A) *Multidrug and toxic compound extrusion 1* (*MATE1*) locus in maize. A 30-kb region containing transposable elements and the *MATE1* gene is triplicated in tandem. The filled red boxes represent each copy of *MATE1.* One copy contains an additional unique gypsy retrotransposon (filled teal). The outlined boxes represent other classes of retrotransposons that are part of the duplicated region. (B) The *shattering1* (*Sh1*) locus in *Oryza barthii* (*ObSh1*) and *Oryza glaberrima* (*OgSh1*). A 45-kb region including the *Sh1* gene (blue), a YABBY transcription factor, and three additional genes (purple) is deleted in domesticated *O. glaberrima* relative to *O. barthii*. This deletion is polymorphic in domesticated populations. (C) Three haplotypes of *Sh1* locus (7758 bp) in nonshattering *Sorghum bicolor* (Sb) relative to the wild, shattering, *Sorghum virgatum* sequence. From top to bottom: a 2.2-kb deletion including two exons, a SNP polymorphism at a splice site, and four SNP variants, two upstream of the transcription start site and two in an intron. Each of these haplotypes is present in nonshattering domesticated species, indicating that CNV is one of multiple mutations that may be causing the loss-of-function trait. Adapted from [67] (A), [102] (B), and [52] (C).

(Figure 2C). In sorghum (*Sorghum bicolor*), both SNP and deletion haplotypes are observed at the *sh1* locus and cause the loss of the shattering phenotype [52] (Figure 2C). The *sh1* deletion CNV remains polymorphic in sorghum, and suggests the multiple domestications of this species [52]. A similar phenomenon is also seen in *O. glaberrima*, where deletion of *sh1* is polymorphic and is found in coastal regions of West Africa (Figure 2B) [107]. This deletion CNV confers the nonshattering phenotype, but does so (possibly in an additive function) with a nonsense allele of the *sh4* gene [107].

A similar example is observed in the transition from a vine growth habit to determinate, compact growth, and earlier flowering associated with domestication of the bean species *Phaseolus vulgaris* [53]. Common beans were domesticated multiple times from multiple gene pools; a comparative analysis of the *PvTFL1y* gene among different cultivars showed that LOF *PvTFL1y* alleles were generated from insertion/deletion (indel) mutations, SNPs, and a large whole-gene deletion [53]. These examples highlight cases of parallel evolution and multiple origins of a

domesticated species that are illuminated by considering CNV mutations. They also demonstrate the role of CNV deletions in generating adaptive LOF alleles during domestication.

CNVs in Postdomestication Diversification

Diversification traits arose following the initial domestication process, as domesticated species moved out of their centers of origin and adapted to different environments, or as desirable traits were selected by specific human cultures. These traits may be consciously selected and are characterized by high selection coefficients and extreme phenotypes [1]. Diversification traits are not always easy to distinguish from domestication traits, and some have argued that these should be considered as part of a spectrum of domestication traits [1].

Various CNVs have been implicated in genes that control specific traits associated with breeds or varieties, and there are clear examples of CNVs found in diversification genes in both domesticated plants and animals. Unlike in domestication genes, where examples arise primarily from deletion CNVs, there appears to be a greater role for gene amplifications in the evolution of diversification traits; indeed, in animals, most known domestication CNVs are duplications underlying intraspecific diversification. One example is the agouti signaling protein (*ASIP*) gene, duplications of which are associated with white coat color in both sheep (*Ovis aries*) and goats (*Capra aegagrus*), suggesting that this phenotype was selected for independently during the diversification of both species [54,55]. Other examples include the ridgeback phenotype in dogs (*Canis familiaris*), caused by a 133-kb duplication containing multiple fibroblast growth factor (*FGF*) loci and the KIT locus, which controls coat color in cattle and pigs [56–58].

Adaptation to new environments was a critical element of postdomestication diversification as humans migrated and spread domesticated species outside their original geographic ranges [3]. In both domesticated and wild plant species, genes affected by CNVs are associated with abiotic stress and disease response [29–31,33–35,37,59–61]. Variation in stress response and disease phenotypes through CNVs could confer preadaptation to new environments during the spread of domesticates and, thus, enable a rapid response to a changing environment.

CNV-associated environmental adaptation is linked to the spread of wheat (*Triticum aestivum*) in Europe during the Neolithic. The *Photoperiod-B1* (*Ppd-B1*) and *Vernalization-A1* (*Vrn-A1*) genes in *T. aestivum* control daylength and **vernalization** requirements for flowering [62]. The spread of wheat cultivation into northern Europe and higher latitudes was associated with a higher copy number of *Vrn-A1*, resulting in an increased vernalization requirement, while additional copies of *Ppd-B1* are associated with earlier flowering times [63]. Different gene copy numbers of both *Ppd-B1* and *Vrn-A1* are present in geographically structured patterns and demonstrate the role of gene duplication in adaptation to new environments as wheat spread to northern latitudes [63].

De novo duplications can be immediately adaptive by increasing gene products [23,64,65]. For example, an amplification of *Bot1*, a boron efflux transporter, confers boron-toxicity tolerance in barley (*Hordeum vulgare*) [66]. In maize, *MATE1* is triplicated in some accessions and increased expression confers aluminum tolerance [67] (Figure 2A). Adaptive increases in expression can also be caused by CNV in regulatory sequences. An example is the amplification of a 17.1-kb sequence upstream of the *GL7* locus in Asian rice, which increases *GL7* gene expression and, thus, seed size [68]. Gene duplications can also result in altered patterns of gene expression rather than simply increasing expression [65]. An example is the genetic variant that results in the pod-corn maize variety, known as *Tunicate* maize. In *Tunicate,* the *ZMM19* MADS-box gene is duplicated and mutated in a *cis*-regulatory region, which causes ectopic gene expression in the inflorescence and creates the pod-corn phenotype [69].

Disease resistance is another important aspect of localized adaptation associated with CNVs in domesticated species. R-genes are often organized in **tandem arrays** and are frequently duplicated and deleted, a process thought to contribute to the evolution of new disease-resistance alleles [9]. The *Rp1* locus in maize demonstrates this phenomenon: different haplotypes of the locus contain different sets and copy numbers of *Rp1* genes, with some haplotypes conferring increased resistance to leaf rust [70].

PAV, another term associated with CNVs in intraspecies diversification, is often discussed using the framework of the **pan-genome** concept [47,71–73].This concept highlights the finding that individuals from a single species may have different genome sizes and genetic compositions. All individuals within a species may have a core set of shared genes, but outside of this core pan-genome gene set, a distributed set of genes are shared by only some individuals and are akin to insertion/deletion genic CNVs. Gene ontology analyses of the pan-genome in plants consistently find that these CNVs are enriched for abiotic stress and disease-response genes [71–73]. The *Sub1* locus in rice, for example, was defined as a PAV gene responsible for submergence tolerance in some rice cultivars [74]. Flooding-tolerant varieties have the *Sub1A* gene in addition to *Sub1B* and *Sub1C* genes [74], while susceptible varieties lack this locus.

Evidence for Selection on CNVs

During domestication and diversification, CNVs underlying domestication and diversification traits increase in frequency in the population in response to selection, and genomic signatures for selection can sometimes be detected associated with these CNVs. Detecting evidence for selection is a complex problem (reviewed in [75]) and, moreover, current CNV detection methods from short read-sequencing data cannot accurately detect CNV boundaries, making it difficult to infer selection based on linkage. Despite these challenges, there is evidence for CNVs under selection in domesticated species. The *Rhg1* locus in soybean, for example, is a 31.2-kb amplification sequence containing four genes that could be under selection [76]. Genotypes with increased copy number in the locus exhibit resistance to cyst nematode disease [76]. *Rhg1* shows high linkage disequilibrium and fixation index (**F_{ST}**) between populations based on SNP data in the regions flanking the CNV, suggesting that this variant was the recent target of selection [76].

As in plants, a similar example of selection for gene amplification is found in domesticated dogs, where increased amylase activity allowed dogs to adapt to a starch-based diet associated with humans [77]. The *AMY2B* locus in domesticated dogs varies in amylase copy number (with individuals having up to 30 copies) and also shows a local reduction in heterozygosity and increased genetic differentiation (F_{ST}) of linked SNPs consistent with positive selection [77,78]. Comparisons of the *AMY2B* locus in different dog breeds showed that human local diet predicts *AMY2B* copy number [78]. This suggests that *AMY2B* copy number was under selection following domestication as dogs and humans spread to diverse habitats and developed different regional diets. Although these studies do show the role of selection on CNVs, increased break-point resolution could increase the ability to infer selection and population dynamics of CNVs.

Concluding Remarks and Future Perspectives

CNVs represent a class of mutations that have key roles in the evolution of domesticated species, but whose impact is not yet fully appreciated. In examining the literature on the genetic basis for variation and evolution in domesticated species, we find that CNVs are primarily associated with postdomestication traits, and may have enormous potential as a source of

Outstanding Questions

To what extent are CNVs inherited from ancestral populations of crop wild relatives? Does the process of domestication favor an increase in the number of CNVs or CNV formation? What is the rate of CNV formation among plants?

How do CNVs contribute to complex domestication traits?

How can CNVs be more accurately identified within populations? What is the relative abundance of CNVs of different sizes within populations?

To what extent are CNVs recurrent? Can complex multiallelic CNVs be distinguished and assigned mechanistic and population origin?

useful traits when crop wild relatives and landraces are used as a source of genetic diversity to improve cultivars.

As sequencing technologies and bioinformatics methods improve our ability to accurately detect CNVs, the opportunity to dissect precise numbers and locations of CNVs will help further elucidate mechanisms of their formation and their evolutionary trajectories. In the future, CNV analyses could be integrated with systems approaches to gene regulatory networks and could show whether genic CNVs are functionally constrained within gene interaction networks. Combining CNV variation with systems biology experimental approaches could yield greater insight into the functional roles of these variants.

Agriculture is an exercise in practical evolutionary genetics, and there is a need to evolve new varieties and breeds that can continue to ensure a stable coevolutionary interaction between humans and their domesticated partners. An appreciation of the role of CNVs may help in targeted modification of crop and livestock species to ensure the future of this coevolutionary interaction, which is so vital to human food security (see also Outstanding Questions).

Acknowledgments

We would like to thank Grace Avecilla, Jae Young Choi, and Rod Wing for useful discussions. This work was funded in part by grants from the US National Science Foundation Plant Genome Research Program, the New York University Abu Dhabi Research Institute and the Zegar Family Foundation to M.D.P.

References

1. Meyer, R.S. and Purugganan, M.D. (2013) Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840–852
2. Larson, G. and Fuller, D.Q. (2014) The evolution of animal domestication. *Annu. Rev. Ecol. Evol. Syst.* 45, 115–136
3. Purugganan, M.D. and Fuller, D.Q. (2009) The nature of selection during plant domestication. *Nature* 457, 843–848
4. Larson, G. *et al.* (2014) Current perspectives and the future of domestication studies. *Proc. Natl. Acad. Sci. U. S. A.* 111, 6139–6146
5. Wang, G.-D. *et al.* (2013) Domestication genomics: evidence from animals. *Annu. Rev. Biosci.* 2, 65–84
6. Hastings, P.J. *et al.* (2009) Mechanisms of change in gene copy number. *Nat. Rev. Genet.* 10, 551–564
7. Alkan, C. *et al.* (2011) Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* 12, 363–376
8. Zmienko, A. *et al.* (2016) *Arabidopsis thaliana* population analysis reveals high plasticity of the genomic region spanning MSH2, AT3G18530 and AT3G18535 genes and provides evidence for NAHR-driven recurrent CNV events occurring in this location. *BMC Genomics* 17, 893
9. Panchy, N. *et al.* (2016) Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316
10. Xiao, H. *et al.* (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* 319, 1527–1530
11. Schnable, J.C. *et al.* (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl. Acad. Sci. U. S. A.* 108, 4069–4074
12. Sun, S. *et al.* (2018) Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nat. Genet.* 50, 1289–1295
13. Hastings, P.J. *et al.* (2009) A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet.* 5, e1000327
14. Katju, V. and Bergthorsson, U. (2013) Copy-number changes in evolution: rates, fitness effects and adaptive significance. *Front. Genet.* 4, 273
15. Emerson, J.J. *et al.* (2008) Natural selection shapes genome-wide patterns of copy-number polymorphism in Drosophila melanogaster. *Science* 320, 1629–1631
16. Redon, R. *et al.* (2006) Global variation in copy number in the human genome. *Nature* 444, 444–454
17. Zhang, F. *et al.* (2009) Copy number variation in human health, disease, and evolution. *Annu. Rev. Genomics Hum. Genet.* 10, 451–481
18. Mileyko, Y. *et al.* (2008) Small-scale copy number variation and large-scale changes in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 105, 16659–16664
19. Chiang, C. *et al.* (2017) The impact of structural variation on human gene expression. *Nat. Genet.* 49, 692
20. Schlattl, A. *et al.* (2011) Relating CNVs to transcriptome data at fine resolution: assessment of the effect of variant size, type, and overlap with functional regions. *Genome Res.* 21, 2004–2013
21. Dopman, E.B. and Hartl, D.L. (2007) A portrait of copy-number polymorphism in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* 104, 19920–19925
22. Keel, B.N. *et al.* (2016) Evolutionary and functional features of copy number variation in the cattle Genome1. *Front. Genet.* 7, 207
23. Gresham, D. *et al.* (2008) The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet.* 4, e1000303
24. Debolt, S. (2010) Copy number variation shapes genome diversity in arabidopsis over immediate family generational scales. *Genome Biol. Evol.* 2, 441–453
25. Farslow, J.C. *et al.* (2015) Rapid Increase in frequency of gene copy-number variants during experimental evolution in *Caenorhabditis elegans*. *BMC Genomics* 16, 1–18
26. Riveron, J.M. *et al.* (2013) Directionally selected cytochrome P450 alleles are driving the spread of pyrethroid resistance in the major malaria vector *Anopheles funestus*. *Proc. Natl. Acad. Sci. U. S. A.* 110, 252–257
27. Puinean, A.M. *et al.* (2010) Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*. *PLoS Genet.* 6, e1000999

CellPress
REVIEWS

28. Edi, C.V. *et al.* (2014) CYP6 P450 enzymes and ACE-1 duplication produce extreme and multiple insecticide resistance in the malaria mosquito *Anopheles gambiae*. *PLoS Genet.* 10, e1004236
29. Zhang, Z. *et al.* (2015) Genome-wide mapping of structural variations reveals a copy number variant that determines reproductive morphology in cucumber. *Plant Cell* 27, 1595–1604
30. Mercenaro, L. *et al.* (2017) Sequence polymorphisms and structural variations among four grapevine (*Vitis vinifera* L.) cultivars representing Sardinian agriculture. *Front. Plant Sci.* 8, 1279
31. McHale, L.K. *et al.* (2012) Structural variants in the soybean genome localize to clusters of biotic stress-response genes. *Plant Physiol.* 159, 1295–1308
32. Sun, S. *et al.* (2018) Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nat. Genet.* 50, 1289–1295
33. Swanson-Wagner, R.A. *et al.* (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Res.* 20, 1689–1699
34. Dolatabadian, A. *et al.* (2017) Copy number variation and disease resistance in plants. *Theor. Appl. Genet.* 130, 2479–2490
35. Muñoz-Amatriaín, M. *et al.* (2013) Distribution, functional impact, and origin mechanisms of copy number variation in the barley genome. *Genome Biol.* 14, R58
36. Boocock, J. *et al.* (2015) The distribution and impact of common copy-number variation in the genome of the domesticated apple, *Malus* x *domestica* Borkh. *BMC Genomics* 16, 848
37. González, V.M. *et al.* (2013) High presence/absence gene variability in defense-related gene clusters of *Cucumis melo*. *BMC Genomics* 14, 782
38. Clop, A. *et al.* (2012) Copy number variation in the genomes of domestic animals. *Anim. Genet.* 43, 503–517
39. Upadhyay, M. *et al.* (2017) Distribution and functionality of copy number variation across European cattle populations. *Front. Genet.* 8, 1–12
40. Paudel, Y. *et al.* (2015) Copy number variation in the speciation of pigs: a possible prominent role for olfactory receptors. *BMC Genomics* 16, 330
41. Sakudoh, T. *et al.* (2011) Diversity in copy number and structure of a silkworm morphogenetic gene as a result of domestication. *Genetics* 187, 965–976
42. Yi, G. *et al.* (2014) Genome-wide patterns of copy number variation in the diversified chicken genomes using next-generation sequencing. *BMC Genomics* 15, 962
43. Ghosh, S. *et al.* (2014) Copy number variation in the horse genome. *PLoS Genet.* 10, e1004712
44. Yu, P. *et al.* (2011) Detection of copy number variations in rice using array-based comparative genomic hybridization. *BMC Genomics* 12, 372
45. Lai, J. *et al.* (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat. Genet.* 42, 1027–1030
46. Swanson-Wagner, R.A. *et al.* (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Res.* 20, 1689–1699
47. Zheng, T. *et al.* (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557, 43–49
48. Hou, Y. *et al.* (2011) Genomic characteristics of cattle copy number variations. *BMC Genomics* 12, 127
49. Jenkins, G.M. *et al.* (2016) Copy number variants in the sheep genome detected using multiple approaches. *BMC Genomics* 17, 441
50. Ramirez, O. *et al.* (2014) Analysis of structural diversity in wolf-like canids reveals post-domestication variants. *BMC Genomics* 15, 465
51. Fuller, D.Q. (2007) Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the Old World. *Ann. Bot.* 100, 903–924
52. Lin, Z. *et al.* (2012) Parallel domestication of the Shattering1 genes in cereals. *Nat. Genet.* 44, 720–724
53. Kwak, M. *et al.* (2012) Multiple origins of the determinate growth habit in domesticated common bean (*Phaseolus vulgaris*). *Ann. Bot.* 110, 1573–1580
54. Norris, B.J. and Whan, V.A. (2008) A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. *Genome Res.* 18, 1282–1293
55. Fontanesi, L. *et al.* (2010) Copy number variation and missense mutations of the agouti signaling protein (ASIP) gene in goat breeds with different coat colors. *Cytogenet. Genome Res.* 126, 333–347
56. Salmon Hillbertz, N.H.C. *et al.* (2007) Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. *Nat. Genet.* 39, 1318–1320
57. Pielberg, G. *et al.* (2002) Unexpectedly high allelic diversity at the KIT locus causing dominant white color in the domestic pig. *Genetics* 160, 305–311
58. Brenig, B. *et al.* (2013) Molecular genetics of coat colour variations in White Galloway and White Park cattle. *Anim. Genet.* 44, 450–453
59. Bai, Z. *et al.* (2016) The impact and origin of copy number variations in the *Oryza* species. *BMC Genomics* 17, 1–12
60. Hardigan, M.A. *et al.* (2016) Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated *Solanum tuberosum*. *Plant Cell* 28, 388–405
61. Zmienko, A. *et al.* (2014) Copy number polymorphism in plant genomes. *Theor. Appl. Genet.* 127, 1–18
62. Díaz, A. *et al.* (2012) Copy number variation affecting the photoperiod-B1 and vernalization-A1 genes is associated with altered flowering time in wheat (*Triticum aestivum*). *PLoS One* 7, e33234
63. Würschum, T. *et al.* (2015) Multiply to conquer: copy number variations at Ppd-B1 and Vrn-A1 facilitate global adaptation in wheat. *BMC Genet.* 16, 96
64. Kondrashov, F.A. (2012) Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc. R. Soc. B Biol. Sci.* 279, 5048–5057
65. Innan, H. and Kondrashov, F. (2010) The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* 11, 97–108
66. Sutton, T. *et al.* (2007) Boron-toxicity tolerance in barley arising from efflux transporter amplification. *Science* 318, 1446–1449
67. Maron, L.G. *et al.* (2013) Aluminum tolerance in maize is associated with higher MATE1 gene copy number. *Proc. Natl. Acad. Sci. U. S. A.* 110, 5241–5246
68. Wang, Y. *et al.* (2015) Copy number variation at the GL7 locus contributes to grain size diversity in rice. *Nat. Genet.* 47, 944–948
69. Wingen, L.U. *et al.* (2012) Molecular genetic basis of pod corn (*Tunicate* maize). *Proc. Natl. Acad. Sci.* 109, 7115–7120
70. Chavan, S. *et al.* (2015) Diversity and evolution of Rp1 rust resistance genes in four maize lines. *Theor. Appl. Genet.* 128, 985–998
71. Zhao, Q. *et al.* (2018) Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* 50, 278–284
72. Golicz, A.A. *et al.* (2016) The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat. Commun.* 7, 13390
73. Li, Y. *et al.* (2014) De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat. Biotechnol.* 32, 1045–1052
74. Xu, K. *et al.* (2006) Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* 442, 705–708
75. Vitti, J.J. *et al.* (2013) Detecting natural selection in genomic data. *Annu. Rev. Genet.* 47, 97–120

76. Lee, T.G. *et al.* (2015) Evolution and selection of Rhg1, a copy-number variant nematode-resistance locus. *Mol. Ecol.* 24, 1774–1791

77. Axelsson, E. *et al.* (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495, 360–364

78. Reiter, T. *et al.* (2016) Dietary variation and evolution of gene copy number among dog breeds. *PLoS One* 11, e0148899

79. Beales, J. *et al.* (2007) A Pseudo-Response Regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 115, 721–733

80. Zhu, J. *et al.* (2014) Copy number and haplotype variation at the VRN-A1 and central FR-A2 loci are associated with frost tolerance in hexaploid wheat. *Theor. Appl. Genet.* 127, 1183–1197

81. Li, Y. *et al.* (2012) A tandem segmental duplication (TSD) in green revolution gene Rht-D1b region underlies plant height variation. *New Phytol.* 196, 282–291

82. Duan, P. *et al.* (2017) Natural variation in the promoter of GSE5 contributes to grain size diversity in Rice. *Mol. Plant* 10, 685–694

83. Kretzschmar, T. *et al.* (2015) A trehalose-6-phosphate phosphatase enhances anaerobic germination tolerance in rice. *Nat. Plants* 1, 15124

84. Hattori, Y. *et al.* (2009) The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. *Nature* 460, 1026–1030

85. Gamuyao, R. *et al.* (2012) The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency. *Nature* 488, 535–539

86. Choi, J.Y. *et al.* (2018) The complex geography of domestication of the African rice *Oryza glaberrima*. *bioRxiv* Published online May 18, 2018. http://dx.doi.org/10.1101/321851

87. Cook, D.E. *et al.* (2012) Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. *Science* 338, 1206–1209

88. Hu, Y. *et al.* (2018) Analysis of extreme phenotype bulk copy number variation (XP-CNV) identified the association ofrp1with resistance to Goss's wilt of maize. *Front. Plant Sci.* 9, 110

89. Zuo, W. *et al.* (2015) A maize wall-associated kinase confers quantitative resistance to head smut. *Nat. Genet.* 47, 151–157

90. Nitcher, R. *et al.* (2013) Increased copy number at the HvFT1 locus is associated with accelerated flowering time in barley. *Mol. Genet. Genomics* 288, 261–275

91. Knox, A.K. *et al.* (2010) CBF gene copy number variation at Frost Resistance-2 is associated with levels of freezing tolerance in temperate-climate cereals. *Theor. Appl. Genet.* 121, 21–35

92. Mu, Q. *et al.* (2017) Fruit weight is controlled by Cell Size Regulator encoding a novel protein that is expressed in maturing tomato fruits. *PLoS Genet.* 13, e1006930

93. Giuffra, E. *et al.* (2002) A large duplication associated with dominant white color in pigs originated by homologous recombination between LINE elements flanking KIT. *Mamm. Genome* 13, 569–577

94. Durkin, K. *et al.* (2012) Serial translocation by means of circular intermediates underlies colour sidedness in cattle. *Nature* 482, 81–84

95. Yang, M. *et al.* (2017) Association study and expression analysis of CYP4A11 gene copy number variation in Chinese cattle. *Sci. Rep.* 7, 46599

96. Sharma, A.K. *et al.* (2011) Chondrodysplasia in dogs. *Indian Vet. J.* 88, 68–70

97. Deane-Coe, P.E. *et al.* (2018) Direct-To-Consumer DNA testing of 6,000 dogs reveals 98.6-kb duplication causing blue eyes and heterochromia in Siberian Huskies. *PLoS Genet.* 14, e1007648

98. Wright, D. *et al.* (2009) Copy number variation in intron 1 of SOX5 causes the pea-comb phenotype in chickens. *PLoS Genet.* 5, e1000512

99. Elferink, M.G. *et al.* (2008) Partial duplication of the PRLR and SPEF2 genes at the late feathering locus in chicken. *BMC Genomics* 9, 391

100. Dharmayanthi, A.B. *et al.* (2017) The origin and evolution of fibromelanosis in domesticated chickens: genomic comparison of Indonesian Cemani and Chinese Silkie breeds. *PLoS One* 12, e0173147

101. Dorshorst, B. *et al.* (2015) A genomic duplication is associated with ectopic eomesodermin expression in the embryonic chicken comb and two duplex-comb phenotypes. *PLoS Genet.* 11, e1004947

102. Wang, M. *et al.* (2014) The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nat. Genet.* 46, 982–988

103. Pirooznia, M. *et al.* (2015) Whole-genome CNV analysis: advances in computational approaches. *Front. Genet.* 6, 1–9

104. Zhao, M. *et al.* (2013) Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC Bioinformatics* 14, S1

105. Wang, H. *et al.* (2015) Genome wide distributions and functional characterization of copy number variations between Chinese and Western pigs. *PLoS One* 10, 1–15

106. Dong, J. *et al.* (2016) Analysis of tandem gene copies in maize chromosomal regions reconstructed from long sequence reads. *Proc. Natl. Acad. Sci. U. S. A.* 113, 7949–7956

107. Choi, J.Y. *et al.* (2018) The complex geography of domestication of the African rice *Oryza glaberrima*. *bioRxiv* https://doi.org/10.1101/32185