CKI and CKII mediate the FREQUENCY-dependent phosphorylation of the WHITE COLLAR complex to close the Neurospora circadian negative feedback loop

Qun He,1,2 Joonseok Cha,1 Qiyang He,1,3 Heng-Chi Lee,1 Yuhong Yang,1,4 and Yi Liu1,5

1Department of Physiology, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA; 2State Key Laboratory for Agro-Biotechnology, College of Biological Sciences, China Agricultural University, Beijing 100094, China

The eukaryotic circadian oscillators consist of circadian negative feedback loops. In Neurospora, it was proposed that the FREQUENCY (FRQ) protein promotes the phosphorylation of the WHITE COLLAR (WC) complex, thus inhibiting its activity. The kinase(s) involved in this process is not known. In this study, we show that the disruption of the interaction between FRQ and CK-1a (a casein kinase I homolog) results in the hypophosphorylation of FRQ, WC-1, and WC-2. In the ck-1aL strain, a knock-in mutant that carries a mutation equivalent to that of the Drosophila dbtL mutation, FRQ, WC-1, and WC-2 are hypophosphorylated. The mutant also exhibits ∼32 h circadian rhythms due to the increase of FRQ stability and the significant delay of FRQ progressive phosphorylation. In addition, the levels of WC-1 and WC-2 are low in the ck-1aL strain, indicating that CK-1a is also important for the circadian positive feedback loops. In spite of its low accumulation in the ck-1aL strain, the hypophosphorylated WCC efficiently binds to the C-box within the frq promoter, presumably because it cannot be inactivated through FRQ-mediated phosphorylation. Furthermore, WC-1 and WC-2 are also hypophosphorylated in the ckaRIT strain, which carries the disruption of the catalytic subunit of casein kinase II. In the ckaRIT strain, WCC binding to the C-box is constantly high and cannot be inhibited by FRQ despite high FRQ levels, resulting in high levels of frq RNA. Together, these results suggest that CKI and CKII, in addition to being the FRQ kinases, mediate the FRQ-dependent phosphorylation of WCs, which inhibit their activity and close the circadian negative feedback loop.

[Keywords: Circadian clock; casein kinase I; knock-in; ChIP; DNA binding; degradation]

Supplemental material is available at http://www.genesdev.org.

Received June 27, 2006; revised version accepted July 27, 2006.

The eukaryotic circadian clocks are composed of auto-regulatory circadian negative feedback loops including both positive and negative elements (Dunlap 1999; Reppert and Weaver 2001; Young and Kay 2001; Sehgal 2004). In Neurospora, Drosophila, and mammals, the positive elements are all heterodimeric complexes, consisting of two PER-ARNT-SIM [PAS] domain-containing transcriptional factors that bind to the cis-elements in the promoter of the negative elements to activate their transcription. On the other hand, the negative elements repress their own transcription by inhibiting the activity of the positive elements through their physical interactions. It is unclear how negative elements inhibit the activity of positive elements to close the circadian negative feedback loops. Since the identification of the Drosophila doubletime (dbt) gene, which encodes for a casein kinase I [CKI] homolog, it has become clear that post-translational protein phosphorylation is essential for the function of circadian clocks (Kloss et al. 1998; Price et al. 1998). Despite the evolutionary distance, remarkable conservation of post-translational regulation exists among different eukaryotic systems from fungi to human (see Discussion; Liu 2005; Heintzen and Liu 2006).

In the filamentous fungus Neurospora crassa, the core circadian negative feedback loop consists of four essen-
Haplotype variation in structure and expression of a gene cluster associated with a quantitative trait locus for improved yield in rice

Guangming He,1,6 Xiaojin Luo,2,3,4,5,6 Feng Tian,2,3,4,5 Kegui Li,1 Zuofeng Zhu,2,3,4,5 Wei Su,1 Xiaoyin Qian,1 Yongcai Fu,2,3,4,5 Xiangkun Wang,2,3,4,5 Chuanqing Sun,2,3,4,5,7 and Jinshui Yang1,7

1State Key Laboratory of Genetic Engineering, Institute of Genetics, School of Life Sciences, Fudan University, Shanghai 2000433, China; 2Department of Plant Genetics and Breeding and State Key Laboratory of Agrobiotechnology, China Agricultural University, Beijing 100094, China; 3Key Laboratory of Crop Heterosis and Utilization of Ministry of Education, Beijing 100094, China; 4Beijing Key Laboratory of Crop Genetic Improvement, Beijing 100094, China; 5Key Laboratory of Crop Genetic Improvement and Genome of Ministry of Agriculture, Beijing 100094, China.

By constructing nearly isogenic lines (NILs) that differ only at a single quantitative trait locus (QTL), we fine-mapped the yield-improving QTL qGY2–1 to a 102.9-kb region on rice chromosome 2. Comparison analysis of the genomic sequences in the mapped QTL region between the donor (Dongxiang wild rice, Oryza rufipogon Griff.) and recurrent (Guichao2, Oryza sativa ssp. indica) parents used for the development of NILs identified the haplotypes of a leucine-rich repeat receptor kinase gene cluster, which showed extensive allelic variation. The sequences between genes in the cluster had a very high rate of divergence. More importantly, the genes themselves also differed between two haplotypes: Only 92% identity was observed for one allele, and another allele was found to have completely lost its allelic counterpart in Guichao2. The other six shared genes all showed >98% identity, and four of these exhibited obvious regulatory variation. The same haplotype segments also differed in length (43.9-kb in Guichao2 vs. 52.6-kb in Dongxiang wild rice). Such extensive sequence variation was also observed between orthologous regions of indica (cv. 93–II) and japonica (cv. Nipponbare) subspecies of Oryza sativa. Different rates of sequence divergence within the cluster have resulted in haplotype variability in 13 rice accessions. We also detected allelic expression variation in this gene cluster, in which some genes gave unequal expression of alleles in hybrids. These allelic variations in structure and expression suggest that the leucine-rich repeat receptor kinase gene cluster identified in our study should be a particularly good candidate for the source of the yield QTL.

[Supplemental material is available online at www.genome.org. The sequence data from this study have been submitted to GenBank under accession nos. AY756174 and DQ195081.]
Identification of IgF, a hinge-region-containing Ig class, and IgD in Xenopus tropicalis

Yaofeng Zhao†‡, Qiang Pan-Hammarström†§, Shuyang Yu†, Nancy Wertz¶, Xiaofeng Zhang†, Ning Li*, John E. Butler†, and Lennart Hammarström†‡

†State Key Laboratory for AgroBiotechnology, China Agricultural University, Beijing 100094, People's Republic of China; ‡Division of Clinical Immunology, Karolinska University Hospital, Huddinge, Karolinska Institutet, SE-141 86 Stockholm, Sweden; §Department of Immunology, School of Basic Medical Science, Peking University, Beijing 100083, People's Republic of China; ¶Department of Microbiology, University of Iowa, Iowa City, IA 52242; and †Centre for Structural Biochemistry, Karolinska Institutet, SE-141 57 Stockholm, Sweden

*State Key Laboratory for AgroBiotechnology, China Agricultural University, Beijing 100094, People's Republic of China; †Division of Clinical Immunology, Karolinska Institutet, SE-141 86 Stockholm, Sweden; ‡Department of Immunology, School of Basic Medical Science, Peking University, Beijing 100083, People's Republic of China; §Department of Microbiology, University of Iowa, Iowa City, IA 52242; and †Centre for Structural Biochemistry, Karolinska Institutet, SE-141 57 Stockholm, Sweden

Edited by Max D. Cooper, University of Alabama, Birmingham, AL, and approved June 12, 2006 (received for review January 11, 2006)

Only three Ig isotypes, IgM, IgX, and IgY, were previously known in amphibians. Here, we describe a heavy-chain isotype in Xenopus tropicalis, IgF (encoded by C_A), with only two constant region domains. IgF is similar to amphibian IgY in sequence, but the gene contains a hinge exon, making it the earliest example, in evolution, of an Ig isotype with a separately encoded genetic hinge. We also characterized a gene for the heavy chain of IgD, located immediately 3′ of C_A, that shares features with the C_D gene in fish and mammals. The latter gene contains eight constant-region-encoding exons and, unlike the chimeric splicing of mammalian IgD, is encoded by a single, continuous gene. The IgF locus in X. tropicalis shows a 5′ V_H-D-J_H-C_A-C_D-C_Y-C_C-3′ organization, suggesting that the mammalian and amphibian Ig heavy-chain loci share a common ancestor.

Immunoglobulins (Igs) are essential components of adaptive immunity and are produced only in gnathostomes such as mammals, birds, reptiles, amphibians, and jawed fish (1, 2). Most mammals express five classes of Igs, IgM, IgD, IgG, IgA, and IgE, each endowed with distinct biological effector functions. Mammalian IgM and IgE heavy chains are composed of four ~110-aa constant region domains encoded by separate exons, presumably arising from gene duplication during evolution (3). IgD and IgG contain only three domains (rodent IgD contains only two constant region domains) but also encompass a short exon-encoded hinge (genetic hinge) (4–7). IgA is also a three-domain molecule, with a functional hinge encoded by the 5′ end of the heavy-chain constant region domain (C_H1) 2 (C_H2) exon (8, 9). The hinge regions contain one or more cysteines that are used to bridge the two heavy chains to form an H_{2}L_{2} antibody structure. Hinge regions are also rich in proline, which confers conformational flexibility that allows waving, rotation of Fab arms, and wagging of the Fc fragment, thus facilitating antigen binding and triggering of effector functions (10, 11). Hinge segments have previously been observed only in mammalian Igs; however, Savan et al. (12) recently identified a heavy-chain isotype in fugu fish that contains a hinge region encoded within its C_H2 exon, similar to the hinge of mammalian IgA (8, 9). This putative hinge contains five repeats of VKPT but lacks a cysteine residue to connect the two heavy chains (12).

It is generally believed that mammalian Igs arose from ancestral Igs of lower vertebrates. IgM is found in all vertebrates (13–18); however, the phylogenetic origin of the remaining mammalian Igs is less well established, although Igs referred to as IgA/IgX and IgY have been reported in birds, reptiles, and amphibians (1, 2, 19–22). Cartilaginous fish and lungfish express IgM, IgNAR, and/or IgW/IgX, containing either two or more than four constant region domains (1, 15, 23). Bony fish express three heavy-chain isotypes, IgM, IgD, and IgZ/IgT (24, 25). IgD is found in most mammals but not in birds, amphibians, or reptiles, whereas multidomain encoding C_A genes have been described in bony fish (1, 22, 26, 27). Thus, a traditional phylogenetic pathway connecting IgD in bony fish and IgD in mammals is missing.

It has long been thought that there are only three Ig classes, IgM, IgA/IgX, and IgY, in lower vertebrates, including birds, reptiles, and amphibians (1, 2, 14, 19–22). The heavy chains that have been characterized to date in these species contain four constant region domains (except for a truncated IgY containing two domains in some species) but no hinge (28). cDNAs encoding the heavy chains of IgM, IgX, and IgY have all been cloned previously in Xenopus laevis (14, 19, 20). IgX has been considered to be an analogue of mammalian IgA because a large number of IgX-positive B cells are located in the gut epithelium (29). IgX is structurally distinct from mammalian IgA but is similar to chicken IgA (19, 30). IgY, also consisting of four constant region domains, is found in a variety of birds, amphibians, and reptiles (31) and is regarded as a functional homologue of IgG and the progenitor of both mammalian IgG and IgE (31).

In the last two decades, molecular approaches have facilitated the investigation of the genomes in a variety of species. IgZ was recently discovered in zebrafish (24). The recent assembly of the Xenopus tropicalis genome sequence allowed us to perform a search for additional Ig heavy-chain constant region genes in an amphibian.

Results

Identification of the Genomic Sequence Encoding IgM (μ), IgX (χ), and IgY (γ) in X. tropicalis. The X. tropicalis genome is available in the X. tropicalis Genome Sequencing Project database at the Sanger Institute (www.sanger.ac.uk/Projects/X.tropicalis/). By using the published Ig sequences in X. laevis as templates, the C_H1, C_H2, and C_Y genes in X. tropicalis were all identified in an assembled scaffold (Scaffold.928) where the C_H1 is located ~40 kb downstream of the C_Y. The constant regions were all deduced on the basis of both genomic sequences and EST clones. The organization of the genes is presented in Fig. 1. The amino acid sequences of all three classes displayed a high degree of divergence between X. tropicalis and X. laevis (Figs. 6–8, which are published as supporting information on the PNAS web site), with sequence similarities of only 74.3%, 82.1%, and 75.0% for IgM, IgX, and IgY, respectively. There is a cysteine located at the carboxy-terminal end of the IgX of X. tropicalis that may be used for binding to the J chain. This cysteine is absent in the IgX sequence of X. laevis (Fig. 7) (19).

Identification of an IgD-Encoding Gene (C_A) in X. tropicalis. The long stretch (~40 kb) of intervening DNA between the C_Y and C_A genes
Arabidopsis MICROTUBULE-ASSOCIATED PROTEIN18 Functions in Directional Cell Growth by Destabilizing Cortical Microtubules

Xia Wang, Lei Zhu, Baoquan Liu, Che Wang, Lifeng Jin, Qian Zhao, and Ming Yuan

INTRODUCTION

Microtubule-associated proteins (MAPs) play critical roles in controlling microtubule (MT) dynamics and organization and hence are involved in the regulation of cell expansion (Lloyd and Chan, 2002; Hussey et al., 2002; Wasteneys and Galway, 2003; Mathur, 2004; Sedbrook, 2004; Smith and Oppenheimer, 2005). Mutations in proteins interacting with MTs result in abnormal plant development and plant cell morphogenesis due to disruption of MT organization. For instance, mutations in MOR1/GEM1, which stabilizes cortical MTs, result in organ twisting and isotropic cell expansion in roots as a result of defects in MT organization and cytokinesis (Twell et al., 2001; Whittington et al., 2001). Mutation of Arabidopsis thaliana SKU6/SPIRAL1, which encodes a plus end–localized MT-interacting protein, causes right-handed axial twisting in roots, etiolated hypocotyls, leaf petioles, and strongly right-skewed root growth on inclined agar media (Sedbrook et al., 2004). The radial expansion defect is also observed in katanin mutants (Burk and Ye, 2002; Webb et al., 2002). RIC1, a novel MT binding protein interacting with ROP GTPase, promotes the organization of cortical MTs locally to inhibit outgrowth of the Arabidopsis pavement cells (Fu et al., 2005). Also, mutations in ATMAP65-1 and ATMAP65-3/PLEIADE result in expanded short root phenotypes resulting from defective cytokinesis (Muller et al., 2004; Smertenko et al., 2004).

Several MT binding domains in interacting proteins have been described. One domain is the repetitive K-K-E-E and K-K-E-I/V motifs, which were first identified in a neural MT-associated protein, MAP1B from mouse. This has no structural relationship with the MT binding domains of kinesin, MAP2, or Tau (Noble et al., 1989). Similar repetitive motifs have also been identified in plants (Figure 1). For instance, a pollen-specific expressed protein, SB401 from Solanum berthaultii, contains six repetitive domains of the sequence V-V-E-K-K-N/E-E (Liu et al., 1997). Another pollen-specific gene, ST901, isolated from a genomic library of the potato diploid species Solanum tuberosum cv Desiree, encodes a hydrophilic protein of 217 amino acid residues, which contains five imperfect repeated motifs of V-V-E-K-K-N/E-E (Hao et al., 2006). Other proteins, such as the pollen-specific Lys-rich protein SBgLR from S. tuberosum (Lang et al., 2004) and TSB from Solanum lycopersicum (Zhao et al., 2004), also contain such repetitive motifs. However, there is no evidence that these are truly plant MAPs.

Here, we report the identification of a plant MAP, MAP18 from Arabidopsis, which was identified by an Arabidopsis genome BLAST search using the sequence of V-V-E-K-K-N/E-E. MAP18 destabilizes MTs and plays an important role in the regulation of MT organization to determine plant directional cell growth.

RESULTS

Identification of the MAP18 Gene and Purification of the Recombinant Protein

A BLAST search of the Arabidopsis genome sequence identified a gene (At5g44610) located on chromosome 5, encoding a protein with unknown function and containing seven repeated motifs of V-E-E-K-K. The full-length cDNA sequence (CDS) encodes a predicted polypeptide of 168 amino acid residues, with an
Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription

Zhuo Du,1 Yiqiang Zhao,1 and Ning Li2

State Key Laboratory for Agrobiotechnology, China Agricultural University, Beijing, 10094, People’s Republic of China

G-quadruplex or G4 DNA, a four-stranded DNA structure formed in G-rich sequences, has been hypothesized to be a structural motif involved in gene regulation. In this study, we examined the regulatory role of potential G4 DNA motifs (PG4Ms) located in the putative transcriptional regulatory region (TRR, −500 to +500) of genes across the human genome. We found that PG4Ms in the 500-bp region downstream of the annotated transcription start site (TSS; PG4M1500) are associated with gene expression. Generally, PG4M1500-positive genes are expressed at higher levels than PG4M1500-negative genes, and an increased number of PG4M1500 provides a cumulative effect. This observation was validated by controlling for attributes, including gene family, function, and promoter similarity. We also observed an asymmetric pattern of PG4M1500 distribution between strands, whereby the frequency of PG4M1500 in the coding strand is generally higher than that in the template strand. Further analysis showed that the presence of PG4M1500 and its strand asymmetry are associated with significant enrichment of RNApol II at the putative TRR. On the basis of these results, we propose a model of G4 DNA-mediated stimulation of transcription with the hypothesis that PG4M1500 contributes to gene transcription by maintaining the DNA in an open conformation, while the asymmetric distribution of PG4M1500 considerably reduces the probability of blocking the progression of the RNA polymerase complex on the template strand. Our findings provide a comprehensive view of the regulatory function of G4 DNA in gene transcription.

[Supplemental material is available online at www.genome.org.]

Genomic DNA predominantly exists in the double-stranded conformation throughout most of the cell cycle; however, certain guanine-rich sequences can fold spontaneously into a four-stranded DNA structure known as a G-quadruplex or G4 DNA. The structure of G4 DNA, which comprises stacked G-tetrads, a square planar arrangement of four guanine bases stabilized by Hoogsteen GG pairing, is extremely stable under physiological conditions (Gellert et al. 1962; Guschlbauer et al. 1990; Han and Hurley 2000; Keniry 2000; Shafer and Smirnov 2000; Simonsson 2001; Burge et al. 2006).

Although relatively little is known about the detailed molecular mechanism by which G4 DNA influences genome function, local DNA structure alternative to the double-stranded conformation might provide regulatory motifs important for gene regulation. Recent studies have shown that the G4 DNA structures formed in the regulatory regions can regulate gene expression (Han and Hurley 2000; Dexeheimer et al. 2006; Maizels 2006; Rawal et al. 2006; Fry 2007). A well-known example is the repressive effect of G4 DNA on transcription of the human MYC gene. The transcriptional activity of the MYC gene is reduced considerably when a parallel G4 DNA, formed in the nucleosome-hypersensitive element III, upstream of the P1 promoter, is stabilized by the G4 ligand TMPyP4. In comparison, a G4 DNA-disrupting mutation caused a threefold increase in the basal promoter activity (Grand et al. 2002; Siddiqui-Jain et al. 2002; Seensamy et al. 2004; Ambiru et al. 2005). Similarly, a reporter assay indicated that stabilization of G4 DNA in the nucleosome-hypersensitive poly purine–polypyrimidine element of the KRAS promoter caused an 80% decrease in transcriptional activity (Cogoi and Xodo 2006). In contrast, several studies have shown a stimulatory role of G4 DNA motifs in gene expression. For example, biochemical and biophysical analyses have indicated that the formation of G4 DNAs in the regulatory regions of the chicken beta-globin and human insulin genes activates transcription through the binding of the G4 DNA-specific proteins (Lewis et al. 1988; Clark et al. 1990; Kennedy and Rutter 1992; Catastini et al. 1996; Lew et al. 2000; Dexeheimer et al. 2006). The G-rich nontemplate strand of rDNA has also been reported to be capable of forming G4 DNA, and this structure has been hypothesized to contribute to the high efficiency of rRNA transcription by interacting with a G4 DNA-specific binding protein, nucleolin (Hanakahi et al. 1999; Maizels 2006).

In addition to these sporadic examples, subsequent studies identified the existence of G4 DNA-forming sequences in the regulatory regions of many other genes. Particularly, G4 DNA motifs have been identified in the promoters of KIT (Rankin et al. 2005; Fernandes et al. 2006), HIF1A (De Armond et al. 2005), VEGFA (Sun et al. 2005), BCL2 (Dai et al. 2006), RBP1 (Xu and Sugiyama 2006), and several muscle-specific genes (Yafe et al. 2005). Genome-wide searches have revealed that potential G4 DNA motifs (PG4Ms) are highly prevalent in the genome and are overrepresented in promoters (Huppert and Balsubramanian 2005, 2007; Todd et al. 2005; Rawal et al. 2006; Du et al. 2007). Moreover, a comparative analysis of PG4Ms across animal species has shown that they are strongly enriched in the transcriptional regulatory region (TRR, defined as 500 bp upstream and downstream of the transcription start site) in warm-blooded animals (Zhao et al. 2007). These combined findings lead to the hypothesis that the G4 DNA structures may be common regulatory elements that play roles in transcriptional regulation through structural transitions in DNA.
Specificity of ARGONAUTE7-miR390 Interaction and Dual Functionality in TAS3 Trans-Acting siRNA Formation

Taiowa A. Montgomery,1,2 Miya D. Howell,2,3 Josh T. Cuperus,1,2 Dawei Li,4 Jesse E. Hansen,2 Amanda L. Alexander,2 Elisabeth J. Chapman,1,2,5 Noah Fahlgren,1,2 Edwards Allen,2,3,6 and James C. Carrington2,3,*

1Molecular and Cellular Biology Program
2Department of Botany and Plant Pathology
3Center for Genome Research and Biocomputing
Oregon State University, Corvallis, OR 97331, USA
4State Key Laboratory for Agro-Biotechnology, China Agricultural University, Beijing 100094, China
5Present address: Department of Biology, Indiana University, Bloomington, IN 47405, USA.
6Present address: Monsanto Company, Chesterfield, MO 63017, USA.
*Correspondence: carrington@cgrb.oregonstate.edu
DOI 10.1016/j.cell.2008.02.033

SUMMARY

Trans-acting siRNA form through a refined RNAi mechanism in plants. miRNA-guided cleavage triggers entry of precursor transcripts into an RNA-DEPENDENT RNA POLYMERASE6 pathway, and sets the register for phased tasiRNA formation by DICER-LIKE4. Here, we show that miR390-ARGONAUTE7 complexes function in distinct cleavage or noncleavage modes at two target sites in TAS3a transcripts. The AGO7 cleavage, but not the noncleavage, function could be provided by AGO1, the dominant miRNA-associated AGO, but only when AGO1 was guided to a modified target site through an alternate miRNA. AGO7 was highly selective for interaction with miR390, and miR390 in turn was excluded from association with AGO1 due entirely to an incompatible 5’ adenosine. Analysis of AGO1, AGO2, and AGO7 revealed a potent 5’ nucleotide discrimination function for some, although not all, ARGONAUTEs. miR390 and AGO7, therefore, evolved as a highly specific miRNA guide/effector protein pair to function at two distinct tasiRNA biogenesis steps.

INTRODUCTION

miRNA and tasiRNA are distinct classes of small RNAs that guide silencing of target RNAs through cleavage or nondegradative repression mechanisms (Chapman and Carrington, 2007). miRNAs arise from transcripts that adopt imperfect, self-complementary foldback structures, whereas tasiRNAs arise from a refined adaptation of the RNAi pathway. TAS transcripts are first processed by miRNA-guided cleavage, which forms a discrete 5’ or 3’ end, and then transcribed by RNA-DEPENDENT RNA POLYMERASE6 (RDR6). The resulting dsRNA is processed into siRNA duplexes in end-dependent, 21 nucleotide steps by DICER-LIKE4 (DCL4) (Allen et al., 2005; Gasiocchi et al., 2005; Peragine et al., 2004; Vazquez et al., 2004; Xie et al., 2005; Yoshikawa et al., 2005). Effector complex formation involves strand separation and selective association with an ARGONAUTE (AGO) protein.

Arabidopsis contains four characterized TAS gene families. TAS1, TAS2, and TAS4 tasiRNA biogenesis initiates with miR173- (TAS1 and TAS2) or miR828-guided (TAS4) cleavage on the 5’ side of the tasiRNA-generating region, while TAS3 tasiRNAs form by miR390-guided cleavage on the 3’ side. miR390 also interacts in a noncleavage mode with a second site near the 5’ end (Axtell et al., 2006; Howell et al., 2007). Features of the TAS3 family, including targeting by miR390, are highly conserved in land plants (Allen et al., 2005; Axtell et al., 2006, 2007; Talmor-Neiman et al., 2006). TAS3 tasiRNAs, but not those from TAS1 or TAS2, are dependent on a specialized ARGONAUTE, AGO7 (also called ZIP) (Adenot et al., 2006; Fahlgren et al., 2006; Hunter et al., 2006). TAS3 tasiRNAs target mRNAs encoding several AUXIN RESPONSE FACTORS (ARF3 and ARF4), negative regulation of which is necessary for proper developmental timing and lateral organ development along the adaxial-abaxial axis (Adenot et al., 2006; Fahlgren et al., 2006; Garcia et al., 2006; Hunter et al., 2006).

The mechanisms for recognition and routing of transcripts through the tasiRNA or RDR6/DCL4-dependent pathway are not well understood. Axtell et al. (2006) proposed a two-hit trigger mechanism, in which transcripts with two or more small RNA target sites are preferentially routed into the RDR6/DCL4 pathway. This explains some aspects of TAS3 tasiRNA formation and the routing of several known, multiply targeted transcripts (Axtell et al., 2006; Chen et al., 2007; Howell et al., 2007), although not necessarily TAS1, TAS2, and TAS4 tasiRNA biogenesis. How the factors associated with miR828, miR173, and miR390 function to provide routing information remains an unresolved problem. In this paper, we show that miR390 is uniquely adapted to initiate TAS3 tasiRNA biogenesis due to
Contrasting Mode of Evolution at a Coat Color Locus in Wild and Domestic Pigs

Meiying Fang1,2, Greger Larson3*, Helena Soares Ribeiro1, Ning Li4, Leif Andersson1,3*

1 Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden, 2 College of Animal Science and Technology, China Agricultural University, Beijing, China, 3 Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden, 4 State Key Laboratory for Agrobiotechnology, China Agricultural University, Beijing, China

**Abstract**

Despite having only begun ~10,000 years ago, the process of domestication has resulted in a degree of phenotypic variation within individual species normally associated with much deeper evolutionary time scales. Though many variable traits found in domestic animals are the result of relatively recent human-mediated selection, uncertainty remains as to whether the modern ubiquity of long-standing variable traits such as coat color results from selection or drift, and whether the underlying alleles were present in the wild ancestor or appeared after domestication began. Here, through an investigation of sequence diversity at the porcine melanocortin receptor 1 (MC1R) locus, we provide evidence that wild and domestic pig (Sus scrofa) haplotypes from China and Europe are the result of strikingly different selection pressures, and that coat color variation is the result of intentional selection for alleles that appeared after the advent of domestication. Asian and European wild boar (evolutionarily distinct subspecies) differed only by synonymous substitutions, demonstrating that camouflage coat color is maintained by purifying selection. In domestic pigs, however, each of nine unique mutations altered the amino acid sequence thus generating coat color diversity. Most domestic MC1R alleles differed by more than one mutation from the wild-type, implying a long history of strong positive selection for coat color variants, during which time humans have cherry-picked rare mutations that would be quickly eliminated in wild contexts. This pattern demonstrates that coat color phenotypes result from direct human selection and not via a simple relaxation of natural selective pressures.

**Citation:** Fang M, Larson G, Soares Ribeiro H, Li N, Andersson L (2009) Contrasting Mode of Evolution at a Coat Color Locus in Wild and Domestic Pigs. PLoS Genet 5(1): e1000341. doi:10.1371/journal.pgen.1000341

**Editor:** Gregory S. Barsh, Stanford University School of Medicine, United States of America

**Received** April 11, 2008; **Accepted** December 15, 2008; **Published** January 16, 2009

**Copyright:** © 2009 Fang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The study was funded by the European Commission PigBioDiv2 project (QLKS-CT-2002-01059), the Swedish Foundation for Strategic Research, and the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning. GL was supported by an EMBO post-doctoral fellowship.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: Leif.Andersson@imbim.uu.se

¤ Current address: Department of Archaeology, Durham University, Durham, United Kingdom

† These authors contributed equally to this work.

**Introduction**

The sizes, shapes, and colors among domestic animals vary significantly more than that of their wild counterparts, often reflecting variation normally associated with genus or family level divergence [1]. Domestication therefore provides an ideal model to test numerous evolutionary questions including the relationship between molecular and morphological change, how the intensification of the relationship between humans and wild plants and animals have altered both players’ genetic and phenotypic constitutions, and whether changes associated with domestication resulted primarily from a release of natural selection pressure, selection on standing genetic variation present in the wild ancestor, or positive selection on novel mutations that have occurred subsequent to domestication.

Coat color variation in domestic animals is of considerable interest in this respect considering that it can be traced back to at least 5,000 years before present when it was documented by administrative officers who recorded the coat color of livestock during the UR III dynasty in Mesopotamia [2]. Modern domestic animal species display a bewildering diversity in coat color, and the melanocortin receptor 1 (MC1R) locus is most consistently polymorphic, having been previously documented and associated with coat color variation in horses, cattle, foxes, pigs, sheep, dogs, and chickens [3–10].

MC1R is a G protein-coupled receptor that is primarily expressed in melanocytes and plays a key role in melanogenesis by determining the switch between production of red/yellow phaeomelanin and dark eumelanin [11]. The binding of melanocyte stimulating hormone (MSH) to MC1R induces synthesis of eumelanin, whereas in the absence of MC1R signaling, melanocytes produce only phaeomelanin. Loss-of-function mutations are therefore associated with recessive red coat color, whereas dominant black coloring is linked with mutations causing constitutive activation of MC1R signaling.

We have previously described the molecular basis for an allelic series at the classical Extension locus (equivalent to MC1R) in pigs [6,7,12]. The wild-type (E+) allele allows full expression of both phaeomelanin and eumelanin. The dominant black color results from two different mutations, each of which evolved independently in Asia and Europe. The E01 allele is Asian in origin and is associated with an L102P missense mutation, and E02 is European and associated with a D124N substitution. The recessive red allele (e) possesses two missense mutations A164V and A243T, though it
A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus

Xiao Yu, Bo Li, Yanping Fu, Daohong Jiang, Said A. Ghabrial, Guoqing Li, Youliang Peng, Jiatao Xie, Jiasen Cheng, Junbin Huang, and Xianhong Yi

*State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan 430070, Hubei Province, China; †Provincial Key Laboratory of Plant Pathology of Hubei Province, Huazhong Agricultural University, Wuhan 430070, Hubei Province, China; ‡Department of Plant Pathology, University of Kentucky, Lexington, KY 40546-0312; and §State Key Laboratories for Agrobiotechnology, China Agricultural University, Beijing 100193, China

Edited by Bradley I. Hillman, Rutgers University, New Brunswick, NJ, and accepted by the Editorial Board March 29, 2010 (received for review November 25, 2009)

Mycoviruses are viruses that infect fungi and have the potential to control fungal diseases of crops when associated with hypovirulence. Typically, mycoviruses have double-stranded (ds) or single-stranded (ss) RNA genomes. No mycoviruses with DNA genomes have previously been reported. Here, we describe a hypovirulence-associated circular ssDNA mycovirus from the plant pathogenic fungus Sclerotinia sclerotiorum. The genome of this ssDNA virus, named Sclerotinia sclerotiorum hypovirulence-associated DNA virus 1 (SsHADV-1), is 2166 nt, coding for a replication initiation protein (Rep) and a coat protein (CP). Although phylogenetic analysis of Rep showed that SsHADV-1 is related to geminiviruses, it is notably distinct from geminiviruses both in genome organization and particle morphology. Polyethylene glycol-mediated transfection of fungal protoplasts was successful with either purified SsHADV-1 particles or viral DNA isolated directly from infected mycelium. The discovery of an ssDNA mycovirus enhances the potential of exploring fungal viruses as valuable tools for molecular manipulation of fungi and for plant disease control and expands our knowledge of global virus ecology and evolution.

mycovirus | ssDNA virus | geminivirus | Sclerotinia sclerotiorum | hypovirulence

Fungi cause catastrophic diseases in all major crops with considerable impact on human lives. For example, the recent re-emergence of the deadly wheat black stem rust fungus is likely to threaten the world’s breadbaskets (1). Application of chemical fungicides is the major method used for control of fungal diseases of economically important crops, especially when resistant cultivars are lacking. To reduce the dependence on fungicides, highly efficient and environmentally friendly alternative methods to control diseases are desirable. Mycovirus-mediated hypovirulence is a phenomenon in which the virulence of fungal pathogens is reduced or even completely lost as a consequence of virus infection (2). Hypovirulence is thought to play a role in counterbalancing plant diseases in nature (3), and it was used successfully to control chestnut blight (caused by the fungus Cryphonectria parasitica) in Europe (4, 5). The successful utilization of hypoviruses for biological control of the chestnut blight fungus has attracted much interest and led to the discovery of hypovirulent strains in other fungi.

Mycoviruses are known to infect and multiply in all major taxa belonging to the kingdom Fungi. Typically, mycoviruses have either double-stranded (ds) or single-stranded (ss) RNA genomes; viruses with DNA genomes have not been reported to infect fungi. Although Rhizidiomyces virus, which infects Rhizidiomyces sp., was previously reported as a dsDNA viral genome (6), the fungal-like host Rhizidiomyces sp. is in fact a water mold belonging to the kingdom Protista (7). Thus, it is not known whether DNA viruses could naturally multiply in fungi.

Geminiviruses, which only infect plants, have circular ssDNA genome and twinned quasi-icosahedral particles (8). Plant diseases caused by geminiviruses have threatened agricultural productivity in most tropical and subtropical areas in the world (9). Nowadays, due to changes in agricultural practices, as well as the increase in global trade in agricultural products, these diseases have spread to more regions (10, 11). Several possible scenarios for the evolution of geminiviruses have been developed, however, there still are many questions to be answered, and it is very difficult to ascertain how ancient geminiviruses or their ancestors are (12–14).

Sclerotinia sclerotiorum is a notorious plant fungal pathogen that inhabits the soil worldwide. It has a wide host range including more than 450 species and subspecies among 64 genera of plants (15). Diseases caused by S. sclerotiorum cannot be controlled efficiently because no resistant cultivars are available and fungicides are difficult to deliver into canopies and soil. Hypovirulent strains of S. sclerotiorum have been isolated (16, 17), and two other RNA viruses infecting the hypovirulent strain Ep-1PN were identified (18–20). In the present study, we report on a unique geminivirus-related ssDNA mycovirus that confers hypovirulence to S. sclerotiorum and thus holds the potential for use in biological control of fungal diseases of important crops.

Results
Hypovirulence and Associated Traits of Strain DT-8 of S. sclerotiorum

Strain DT-8 grew slowly on potato dextrose agar (PDA) and developed colony morphology similar to that of the hypovirulent strain Ep-1PN (Fig. 1 A and D). The hypovirulence phenotype of strain DT-8 is even more pronounced than that of strain Ep-1PN when inoculated onto Arabidopsis thaliana (Fig. 1 B) or detached leaves of Brassica napus (Fig. S1). Although strain DT-8 was able to produce sclerotia, the time required for sclerotial initiation was about 3–5 days longer than that for the normal strain and the sclerotia were randomly distributed on the colony surface. Furthermore, the size of the sclerotia produced by strain DT-8 was significantly smaller than that of the normal strain (Fig. 1 C).

Strain DT-8 was cured of its hypovirulent phenotype either by lyophil-tip culturing or protoplast regeneration, and the cured isolates of DT-8, such as DT-SVF, regained virulence on plants.
DCAF26, an Adaptor Protein of Cul4-Based E3, Is Essential for DNA Methylation in *Neurospora crassa*

Hui Xu¹, Jiyong Wang¹, Qiwen Hu¹, Yun Quan¹, Huijie Chen¹, Yingqiong Cao¹, Chunbo Li², Ying Wang¹, Qun He¹*

¹State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing, China, ²Biomedical Analysis Center, Tsinghua University, Beijing, China

Abstract

DNA methylation is involved in gene silencing and genome stability in organisms from fungi to mammals. Genetic studies in *Neurospora crassa* previously showed that the CUL4-DDB1 E3 ubiquitin ligase regulates DNA methylation via histone H3K9 trimethylation. However, the substrate-specific adaptors of this ligase that are involved in the process were not known. Here, we show that, among the 16 DDB1- and Cul4-associated factors (DCAFs) encoded in the *N. crassa* genome, three interacted strongly with CUL4-DDB1 complexes. DNA methylation analyses of *dcaf26* knockout mutants revealed that *dcaf26* was required for all of the DNA methylation that we observed. In addition, histone H3K9 trimethylation was also eliminated in *dcaf26* knockout mutants. Based on the finding that DCAF26 associates with DDB1 and the histone methyltransferase DIM-5, we propose that DCAF26 protein is the major adaptor subunit of the CUL4-DDB1-DCAF26 complex, which recruits DIM-5 to DNA regions to initiate H3K9 trimethylation and DNA methylation in *N. crassa*.

Introduction

The CUL4-DDB1 complex, a major class of cullin-RING ubiquitin ligases (CRLs), is evolutionarily conserved from yeasts to humans. Previous studies have indicated that Cul4-DDB1-regulated ubiquitination is linked to multiple processes, such as cell cycle regulation, DNA replication licensing, DNA repair, and gene expression processes [1–3]. In the CRLs, culin associates with substrates via adaptor molecules in the N terminus, and interacts with the E2 enzyme via the RING finger protein Hrt1/ROC1/ Rbx1 in the C terminus [4]. Substrate-specific adaptors, such as the F-box-containing proteins in SCF complexes and the BTB domain-containing proteins in Cul3-based ubiquitin ligases, determine substrate-specific ubiquitination in many biological processes [5]. Although the different structural states of DDB1 may allow it to directly recruit substrates to the Cul4-based E3 platform, studies have demonstrated that ubiquitination of several characterized CUL4-DDB1 substrates requires additional substrate-specific adaptors [6,7].

Recent studies showed that a class of adaptors called DCAFs (DDB1- and Cul4-associated factors) [4] are employed by Cul4-based E3 ligases to identify specific proteins for ubiquitination. Most DCAFs are WD40-containing proteins with relatively conserved “WDXR” motifs that interact with DDB1 protein. However, several DCAF proteins lacking this conserved motif are still able to bind DDB1 in vivo [3,4,8,9].

Among the well-characterized DCAF proteins, mammalian WDR5 and RBBP5 are essential components of the histone methyltransferase complex that methylates histone H3 on lysine 4 (H3K4) [10–12]. Further studies showed that Cul4-DDB1 can interact with WDR5 and RBBP5 and regulate histone H3K4 methylation. Down regulation of each of these genes by siRNA severely reduces the tri- and monomethylation of histone H3K4, but not H3K4 dimethylation [13]. Interestingly, inactivation of Cul4 or DDB1 also causes a significant inhibition of histone H3K9 and H3K27 trimethylation [13]. However, none of these DCAFs were shown to be involved directly in DNA methylation in eukaryotes.

In fission yeast, the Cul4-Rik1 E3 ubiquitin ligase associates with the histone methyltransferase Clr4 on heterochromatic regions to methylate histone H3K9, contributing to heterochromatin assembly and maintenance [14]. The catalytic activity of Cul4 is required for its proper function in heterochromatin formation. This study suggests that the activity of Cul4-based E3 ligase is required for histone H3K9 methylation. In addition, a WD-40-containing protein, Raf1/Dos1/Clr8/Cmc1, is required for histone H3K9 methylation and heterochromatin formation in *S. pombe* [15–18]. Thus, these studies imply that Raf1/Dos1/Clr8/Cmc1 functions as an adaptor protein associated with Cul4-Rik1 complex in *S. pombe*.

We previously demonstrated that Cul4-DDB1 E3 ligase is essential for DNA methylation in *N. crassa* by regulating histone H3K9 trimethylation [19]. These results suggest that Cul4-DDB1 ubiquitin ligase is required for epigenetic control in higher eukaryotes. However, the substrate-specific adaptors of Cul4-DDB1 E3 ligase and the requirement of DCAFs as the substrate adaptors in DNA methylation are unknown in *N. crassa*.
Genome-wide patterns of genetic variation among elite maize inbred lines

Jinheng Lai1,2,7, Ruiqiang Li3,7, Xun Xu3,7, Weiwei Jin2,7, Mingliang Xu2,7, Hainan Zhao1,2, Zhongkai Xiang1,2, Weibin Song1,2, Kai Ying3, Mei Zhang1,2, Yinpeng Jiao1,2, Peixiang Ni3, Jianguo Zhang3, Dong Li3, Xiaosen Guo3, Kaixiong Ye3, Min Jian3, Bo Wang3, Huisong Zheng3, Huiqing Liang3, Xiuqing Zhang3, Shoucai Wang2, Shaojiang Chen2, Jiansheng Li2, Yan Fu4, Nathan M Springer5, Huanming Yang3, Jian Wang3, Jingrui Dai2, Patrick S Schnable4 and Jun Wang3,6

We have resequenced a group of six elite maize inbred lines, including the parents of the most productive commercial hybrid in China. This effort uncovered more than 1,000,000 SNPs, 30,000 indel polymorphisms and 101 low-sequence-diversity chromosomal intervals in the maize genome. We also identified several hundred complete genes that show presence/absence variation among these resequenced lines. We discuss the potential roles of complementation of presence/absence variations and other deleterious mutations in contributing to heterosis. High-density SNP and indel polymorphism markers reported here are expected to be a valuable resource for future genetic studies and the molecular breeding of this important crop.

The maize genome is large and complex1–4. Its genetic variation has been characterized by using molecular markers and by sequencing multiple alleles from selected loci5–8. With the advent of ‘next generation’ sequencing technology, it has become feasible to resequence entire large genomes9,10 and thereby to carry out genome-wide surveys of genetic variation. The sequencing of the inbred B73 maize line9 provides a reference genome that can be used to anchor resequencing data from other maize lines. Here, we analyze the whole-genome resequencing of six elite commercial maize inbred lines.

Inbred lines (Zheng58, 5003, 478, 178, Chang7-2 and Mo17) were selected on the basis of their agronomic importance and genetic relationships. Lines Zheng58, Chang7-2, 178 and Mo17 are all members of a popular heterotic group used in China (Mo17 is also a member of an important heterotic group used in the USA). Zheng58 and Chang7-2 are the parents of the commercial hybrid (ZD958) that is currently the most widely grown in China. Inbred line 178 is the female parent of another hybrid (ND108) that is also widely grown in China. Inbred line 478 is a parent and inbred line 5003 is a grandparent of Zheng58 (Fig. 1).

Resequencing yielded 1.26 billion 75-bp paired-end reads, which comprised 83.7 Gb of high-quality raw data. Sequence reads were aligned to the maize reference genome using SOAP software v2.18 (ref. 10). In total, we achieved an effective depth of ×32.4 coverage, with an average of ×5.4 for each inbred line (Supplementary Table 1).

We used unique reads in non-repeat regions to detect SNPs and indel polymorphisms (IDPs). SNPs were called with SOAPsnp11 using a conservative quality filter pipeline (Online Methods). We identified 1,272,134 SNPs in non-repeat regions, with 468,966 in the 32,540 high-confidence maize genes (the ‘filtered gene set’) and 130,053 SNPs in coding regions. We also identified 30,178 indels ranging from 1 bp to 6 bp in length, of which 571 were in coding regions (Supplementary Table 2). Owing to the inherent relationships between some samples and the characteristics of inbred lines, the

Figure 1 Genetic background of three sequenced inbred lines. (a) Pedigrees of three resequenced inbred lines. The female parent of each cross is listed first. The name of one parent of Zheng58 is not recorded and is termed ‘Inbred X’. Resequenced inbred lines are underlined. (b) Reconstructed recombination events in inbred lines 478 and Zheng58 as they were derived from their parental lines.

1 State Key Lab of Agrobiotechnology, China Agricultural University, Beijing, China. 2National Maize Improvement Center, China Agricultural University, Beijing, China. 3BGI-Shenzhen, Shenzhen, China. 4Center for Plant Genomics, Iowa State University, Ames, Iowa, USA. 5Department of Plant Biology, University of Minnesota, Saint Paul, Minnesota, USA. 6Department of Biology, University of Copenhagen, Copenhagen, Denmark. 7These authors contributed equally to this work. Correspondence should be addressed to Jun Wang (wangj@genomics.cn), J. Lai (lai@cau.edu.cn) or P.S.S. (schnable@iastate.edu).

Received 16 October 2009; accepted 13 September 2010; published online 24 October 2010; doi:10.1038/ng.684
Granulosa Cell Ligand NPPC and Its Receptor NPR2 Maintain Meiotic Arrest in Mouse Oocytes

Meijia Zhang\textsuperscript{1,2}, You-Qiang Su\textsuperscript{2}, Koji Sugiura\textsuperscript{2,3}, Guoliang Xia\textsuperscript{1,*}, and John J. Eppig\textsuperscript{2,*}

\textsuperscript{1}State Key Laboratory for Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing, 100193, P.R. China

\textsuperscript{2}The Jackson Laboratory, Bar Harbor, ME 04609, USA

\textsuperscript{3}Present address: Laboratory of Applied Genetics, Graduate School of Agricultural and Life Science, University of Tokyo, Tokyo, Japan

Abstract

Granulosa cells of mammalian Graafian follicles maintain oocytes in meiotic arrest, which prevents the precocious maturation. We show that mouse mural granulosa cells, which line the follicle wall, express natriuretic peptide precursor type C, \textit{Nppc}, mRNA while cumulus cells surrounding oocytes express mRNA of the NPPC receptor NPR2, a guanylyl cyclase. NPPC elevated cGMP levels in cumulus cells and oocytes and inhibited meiotic resumption in vitro. Meiotic arrest was not sustained in most Graafian follicles of \textit{Nppc} or \textit{Npr2} mutant mice, and meiosis resumed precociously. Oocyte-derived paracrine factors promoted cumulus cell expression of \textit{Npr2} mRNA. Therefore, the granulosa cell ligand NPPC and its receptor NPR2 in cumulus cells prevent precocious meiotic maturation, which is critical for maturation and ovulation synchrony and for normal female fertility.

Meiosis is a germ cell-specific process that reduces the number of chromosomes from the diploid to the haploid number. It begins in human and mouse ovaries during fetal life but meiotic progression becomes arrested for prolonged periods at the diplotene stage of meiotic prophase. Fully-grown mammalian oocytes in Graafian follicles are maintained in meiotic prophase arrest until the preovulatory surge of luteinizing hormone (LH) triggers the resumption of meiosis and ovulation. The mature oocytes (eggs) are then available for fertilization within the oviduct. The somatic cell compartment of Graafian follicles, consisting of mural granulosa cells lining the inside of the follicle wall and cumulus cells surrounding the oocyte, plays a crucial role in maintaining oocyte meiotic arrest in mammals since removal of the oocyte-cumulus cell complex from these follicles results in gonadotropin-independent meiotic resumption in culture (1,2). Cyclic nucleotides cAMP and cGMP are crucial to the maintenance of meiotic arrest. Cyclic AMP is generated within oocytes downstream of GPR3 and GPR12, regulators of Gs proteins controlling adenylyl cyclase (3,4). Inability to sustain oocyte cAMP levels leads to precocious gonadotropin-independent resumption of meiosis, which interrupts the synchrony between oocyte maturation and ovulation and compromises female fertility (3-5). PDE3A, an oocyte-specific phosphodiesterase, becomes activated after the LH-surge to decrease cAMP levels in oocytes and thereby initiates pathways governing meiotic resumption (6). Before the LH-surge, cGMP, originating in granulosa cells of the follicular somatic compartment and transferred to the oocyte via gap junctions, inhibits activity of PDE3A in the oocyte (7,8). Therefore, control of cGMP production by granulosa cells is crucial for maintaining meiotic arrest in fully-grown oocytes.

\textsuperscript{*}To whom correspondence should be addressed. John.Eppig@jax.org; glxia@cau.edu.cn

\textsuperscript{3}Present address: Laboratory of Applied Genetics, Graduate School of Agricultural and Life Science, University of Tokyo, Tokyo, Japan
Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA

Grexer Larson,a,b,1 Ranran Liu,c,1 Xingbo Zhao,f, Jing Yuan,f Dorian Fuller,d Loukas Barton,g,h Keith Dobney,i Qipeng Fan,i Zhiliang Guj, Xiao-Hui Liuj, Yunbing Luoig, Peng Lvj, Leif Anderssonk and Ning Li*c

a State Key Laboratory of Agrobiotechnology, China Agricultural University, Beijing 100193, China; b Department of Archaeology, University of Durham, Durham DH1 3L, United Kingdom; c Research Centre for Archaeological Science, Institute of Archaeology, Chinese Academy of Social Sciences, Beijing 100710, China; d Department of Medical Biochemistry and Microbiology, Uppsala University, SE-751 23, Uppsala, Sweden; University College London, Institute of Archaeology, London, WC1H 0PY, England; f Katmai National Park and Preserve, King Salmon, AK 99613; g Department of Anthropology, University of California, Davis, CA 95616; h Center for Arid Environment and Paleoclimate Research, Lanzhou University, Lanzhou 730000, China; i Department of Archaeology, University of Aberdeen, Aberdeen AB24 3JF, Scotland, United Kingdom; and j Hubei Provincial Institute of Cultural Relics and Archaeology, Wuhan, 430077, China

The establishment of agricultural economies based upon domestic animals began independently in many parts of the world and led to both increases in human population size and the migration of people carrying domestic plants and animals. The precise circumstances of the earliest phases of these events remain mysterious given their antiquity and the fact that subsequent waves of migrants have often replaced the first. Through the use of more than 1,500 modern (including 151 previously uncharacterized specimens) and 18 ancient (representing six East Asian archeological sites) pig (Sus scrofa) DNA sequences sampled across East Asia, we provide evidence for the long-term genetic continuity between modern and ancient Chinese domestic pigs. Although the Chinese case for independent pig domestication is supported by both genetic and archaeological evidence, we discuss five additional (and possibly) independent domestications of indigenous wild boar populations: one in India, three in peninsular Southeast Asia, and one off the coast of Taiwan. Collectively, we refer to these instances as “cryptic domestication,” given the current lack of corroborating archaeological evidence. In addition, we demonstrate the existence of numerous populations of genetically distinct and widespread wild boar populations that have not contributed maternal genetic material to modern domestic stocks. The overall findings provide the most complete picture yet of pig evolution and domestication in East Asia, and generate testable hypotheses regarding the development and spread of early farmers in the Far East.

Asian colonization | mtDNA | phylogeography

The transition from hunting wild animals to stock-raising occurred in many places independently across the globe. In many cases, the combination of herding and agriculture spurred an increase in population size brought about by the advent of sedentary living, which in turn spurred a demographic migration outward from centers of domestication and agricultural origin (1). An understanding of the locations, timing, and processes of domestication are therefore essential to understanding not only the roots of modern civilization, but also the migratory trajectories that have shaped the modern geography of human languages and cultures (2).

The study of domestication has been transformed by molecular data that have provided new insights, not least of which has been the general conclusion that animal domestication occurred more often and in more places than had been suggested by traditional archaeological evidence (3, 4). Recent publications have sought to further explore the domestication of pigs using both archaeological (5–8) and genetic approaches (9–14). These studies confirmed long-suspected separate domestications of different subspecies of European and Asian wild boar, and revealed that wild boar from regions such as Italy (10) and India (10, 14) were also either domesticated or at least incorporated into domestic stock originally derived from regionally differentiated wild populations. A clear correlation between phylogenetic signals and geographic provenance also allows pigs to be important proxies of human dispersal. As such, they have already revealed the movement of domesticated Near Eastern pigs into Neolithic Europe (9), as well as an unexpected (possibly Austronesian) trajectory connecting Southeast Asia to Oceania (11).

Recent studies in East Asia have highlighted the antiquity of the origins of agriculture and the domestication of plants and animals. Among current views are that early agricultural activities practiced by seasonally mobile cultivators focused on millets in northern China were well established along the Yellow River and Inner Mongolia by ~8,000 (cal) B.P. (5, 8, 15, 16), and domestication may have even begun 2,000 years earlier (17). In southern China, available evidence can be interpreted to suggest that it was sedentary hunter-gatherers (18) who first began cultivating rice along the Yangtze about 9,000–8,000 B.P., a process that culminated in the dependence upon domesticated rice agriculture by ~6,000 B.P. (19). Although dogs were likely the earliest domesticated animal in these regions, available zooarchaeological evidence has been interpreted to indicate that domestic pigs were prevalent in both northern and southern China by at least 8,000 B.P. (6, 7, 20). In both regions, however, pigs make up a small percentage of the earliest mammal bone assemblages that are instead dominated by remains of hunted deer (7, 21).

The extent to which pig domestication in each region was independent or connected by diffusion from a single origin remains to be established, although recent research based on complete mitochondrial genomes of East Asian pigs suggests that Chinese wild boar may have been domesticated independently in the Mekong and in the middle to downstream Yangtze regions (13), although the geographic boundaries of these regions were not defined. Current archaeological evidence suggests that once established, intensive, sedentary agriculture (including rice, millet, and pigs) expanded across various regions from Northeast Asia and Central China (see Table 1 for regional definitions) about 6,500–5,000 B.P., and southward from the Yangtze about 5,000–4,000 B.P. during the Late Neolithic (22, 23). The current understanding is that domesticated pigs and rice are also present in Thailand no earlier than ~4,000 B.P. (24), although the evidence for pigs is based upon traditional metrical analyses. A

Author contributions: N.L., X-B.Z., J.Y., G.L., and R-R.L. designed research; R-R.L., G.L., X-B.Z., J.Y., Q-P.F., Z-L.G., X-H.L., Y-B.L., and P.L. performed research; L.A. contributed

Freely available online through the PNAS open access option. Database deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. F601390–F601547).

1G.L. and R.R.L. contributed equally to this work.
2To whom correspondence may be addressed. E-mail: greger.larson@durham.ac.uk or ninglcau@cau.edu.cn.

This article contains supporting information online at www.pnas.org/cgi/content/full/0912264107/DCSupplemental.

Edited by Ofer Bar-Yosef, Harvard University, Cambridge, MA, and approved March 22, 2010 (received for review October 28, 2009)
RNA-mediated viral immunity requires amplification of virus-derived siRNAs in Arabidopsis thaliana

Xian-Bing Wang,a,b Qingfa Wu,b Takao Ito,c Fabrizio Cillo,c Wan-Xiang Li,d Xuemei Chend, Jia-Lin Yu,e and Shou-Wei Dinga,b

aCenter for Plant Cell Biology, Department of Plant Pathology and Microbiology, and bDepartment of Botany and Plant Sciences, University of California, Riverside, CA 92521; cState Key Laboratory of Agro-Biotechnology, China Agricultural University, Beijing 100094, China; and dIstituto di Virologia Vegetale, CNR, Via Amendola 165/A, 70126 Bari, Italy

Edited by Peter Palese, Mount Sinai School of Medicine, New York, NY, and approved November 13, 2009 (received for review April 14, 2009)

In diverse eukaryotic organisms, Dicer-processed, virus-derived small interfering RNAs direct antiviral immunity by RNA silencing or RNA interference. Here we show that in addition to core dicing and slicing components of RNAi, the RNAi-mediated viral immunity in Arabidopsis thaliana requires host RNA-directed RNA polymerase (RDR) 1 or RDR6 to produce viral secondary siRNAs following viral RNA replication-triggered biogenesis of primary siRNAs. We found that the two antiviral RDRs exhibited specificity in targeting the tripartite positive-strand RNA genome of cucumber mosaic virus (CMV). RDR1 preferentially amplified the 5′-terminal siRNAs of each of the three viral genomic RNAs, whereas an increased production of siRNAs targeting the 3′ half of RNA3 detected in rdr1 mutant plants appeared to be RDR6-dependent. However, siRNAs derived from a single-stranded 336-nucleotide satellite RNA of CMV were not amplified by either antiviral RDR, suggesting avoidance of the potent RDR-dependent silencing as a strategy for the molecular parasite of CMV to achieve preferential replication. Our work thus identifies a distinct mechanism for the amplification of immunity effectors, which together with the requirement for the biogenesis of endogenous siRNAs, may play a role in the emergence and expansion of eukaryotic RDRs.

antiviral immunity | cucumber mosaic virus | RNA silencing | RNA-dependent RNA polymerase | secondary small interfering RNA

RNA silencing or RNA interference in fungi, nematodes and plants requires amplification of small interfering RNAs by eukaryotic RNA-directed RNA polymerases (RDRs) (1–3). Plant and fungal RDRs convert transcripts of target genes into dsRNA that is subsequently processed into secondary siRNAs by a Dicer or Dicer-like (DCL) nuclease. In contrast, Caenorhabditis elegans RDRs, such as RRF-1, may directly manufacture secondary siRNAs without dicing a dsRNA precursor (1–3). The genome of Arabidopsis thaliana encodes six RDRs that are grouped into four clusters (1, 2, 4), among which little is known about cluster III, consisting of RDRs 3a, 3b, and 3c. RDR2 and RDR6 are both required for the short-distance spread of transgene silencing and for the perception, but not the production, of the long-distance mobile silencing signal (1, 5–9).

RDR2 is also essential for the biogenesis of the DCL3-dependent 24-nucleotide (nt) repeat-associated siRNAs (rasiRNAs) derived from transposons, retroelements, and other elements, which are the most abundant endogenous small RNAs in A. thaliana (3). Similarly, RDR6 coupled with DCL4 or DCL1 is responsible for the biogenesis of transacting siRNAs (tasiRNAs) and natural antisense siRNAs (nat-siRNAs), which silence expression of their target genes like microRNAs (miRNAs) (1–3).

RNA silencing controls antiviral immunity in fungi, plants, and invertebrates by producing virus-derived siRNAs to be loaded in an Argonaute protein for antiviral silencing (10–12). In A. thaliana, DCL4 and DCL2 produce viral siRNAs against distinct positive (+)-strand RNA viruses in a hierarchical and redundant manner (13–16). RDR1 and RDR6 of plants and RRF-1 of C. elegans have been implicated in antiviral silencing because RDR-defective mutants exhibit enhanced susceptibility to some of the RNA viruses examined (5, 13, 17–25). However, although two recent studies detected RDR-dependent biogenesis of viral siRNAs in A. thaliana, neither observed a consistent effect of secondary siRNAs on antiviral silencing (13, 18). For example, a modestly reduced viral siRNA production observed in rdr1 mutant plants was not associated with an expected increase in virus accumulation (13). Similarly, although virus accumulated to higher levels in both rdr2 rdr6 and rdr1 rdr2 rdr6 plants than in WT plants, a markedly reduced accumulation of viral siRNAs was detected only in rdr1 rdr2 rdr6 plants (18). Thus, it remains to be established if host RDRs regulate virus resistance either directly by amplification of viral siRNAs or indirectly by the activity of the endogenous RDR-dependent siRNAs of the host. In this regard, it should be pointed out that silencing against RNA viruses does not have to involve a host RDR. RNA viruses encode their own RNA-dependent RNA polymerase that synthesizes dsRNA replicative intermediates during viral RNA replication, which are sufficient to induce antiviral silencing in Drosophila melanogaster, an organism that neither encodes an RDR ortholog nor produces secondary siRNAs (26).

Plant and animal viruses encode essential pathogenesis factors to act as viral suppressors of RNA silencing (VSRs) (27). Recent studies have shown that use of VSR-deficient mutant viruses facilitates mapping the genetic requirements of the RNAi-mediated viral immunity (13, 14, 28, 29). Removal of VSRs enhances virus sensitivity to the immunity and thus is essential to reveal some of the induced antiviral silencing events that would otherwise be undetectable in hosts infected by a wild-type virus. In this study, we investigated the role of A. thaliana RDRs in the RNAi-mediated viral immunity by using a mutant of cucumber mosaic virus (CMV) that does not express the VSR protein 2b. CMV contains three positive-strand genomic RNAs and the 2b protein encoded by RNA2 is essential for infection by suppressing antiviral silencing initiated by either DCL4 or DCL2 (13). Our results demonstrate an essential role for the amplification of viral siRNAs by either RDR1 or RDR6 in antiviral silencing. Further analyses, including illumina sequencing of more than 3.5 million viral siRNAs, indicated target specificity of the two antiviral RDRs. The possibility that emergence and expansion of eukaryotic RDRs represent an evolutionary adaptation to virus infection is discussed.

Author contributions: X. B. W., Q. W., J.-L. Y., and S.-W. D. designed research; X. B. W., Q. W., J.-L. Y., and S.-W. D. performed research; F. C. and X. C. contributed new reagents/analytic tools; X. B. W., Q. W., and S.-W. D. analyzed data; and X. B. W. and S.-W. D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

To whom correspondence should be addressed. E-mail: shou-wei.ding@ucr.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0904086107/DCSupplemental.
Role of Individual Subunits of the *Neurospora crassa* CSN Complex in Regulation of Deneddylation and Stability of Cullin Proteins

Jiyong Wang¹, Qiwen Hu¹, Huijie Chen¹, Zhipeng Zhou¹, Weihua Li², Ying Wang¹, Shaojie Li³, Qun He¹*

¹State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing, China, ²Institute of Basic Medical Sciences, National Center of Biomedical Analysis, Beijing, China, ³Key Laboratory of Systematic Mycology and Lichenology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China

Abstract

The Cop9 signalosome (CSN) is an evolutionarily conserved multifunctional complex that controls ubiquitin-dependent protein degradation in eukaryotes. We found seven CSN subunits in *Neurospora crassa* in a previous study, but only one subunit, CSN-2, was functionally characterized. In this study, we created knockout mutants for the remaining individual CSN subunits in *N. crassa*. By phenotypic observation, we found that loss of CSN-1, CSN-2, CSN-4, CSN-5, CSN-6, or CSN-7 resulted in severe defects in growth, conidiation, and circadian rhythm; the defect severity was gene-dependent. Unexpectedly, CSN-3 knockout mutants displayed the same phenotype as wild-type *N. crassa*. Consistent with these phenotypic observations, deneddylation of cullin proteins in *csn-1, csn-2, csn-4, csn-5, csn-6, or csn-7* mutants was dramatically impaired, while deletion of *csn-3* did not cause any alteration in the neddylation/deneddylation state of cullins. We further demonstrated that CSN-1, CSN-2, CSN-4, CSN-5, CSN-6, and CSN-7, but not CSN-3, were essential for maintaining the stability of Cul1 in SCF complexes and Cul3 and BTB proteins in Cul3-BTB E3s, while five of the CSN subunits, but not CSN-3 and CSN-5, were also required for maintaining the stability of SKP-1 in SCF complexes. All seven CSN subunits were necessary for maintaining the stability of Cul4-DDB1 complexes. In addition, CSN-3 was also required for maintaining the stability of the CSN-2 subunit and FWD-1 in the SCF<sup>FWD-1</sup> complex. Together, these results not only provide functional insights into the different roles of individual subunits in the CSN complex, but also establish a functional framework for understanding the multiple functions of the CSN complex in biological processes.

Introduction

The Cop9 signalosome (CSN) is a multiprotein complex that was initially discovered in *Arabidopsis thaliana* as an important regulator of photomorphogenesis, and was later found to participate in a wide range of processes in eukaryotes [1]. The CSN usually contains eight subunits (CSN1–CSN8) in higher eukaryotes, and each CSN subunit has evolutionarily conserved counterparts in the 26S proteasome lid complex and eukaryotic translation initiation factor 3 (eIF3) [2,3]. All known CSNs regulate ubiquitin-dependent protein degradation [4].

The ubiquitin–proteasome system is the major pathway responsible for the degradation of intracellular proteins. In this pathway, proteins targeted for rapid degradation are conjugated to ubiquitin, a small conserved protein with 76 amino acids [5]. The attachment of ubiquitin to its target proteins is mediated by a cascade of enzymatic reactions involving the ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2), and ubiquitin ligase (E3). After recruiting the specific substrate, the ubiquitin ligase (E3) complex bridges the targeted protein and E2-ubiquitin to form a polyubiquitinated protein, which is subsequently degraded by the 26S proteasome [6]. Cullin-RING ubiquitin ligases (CRLs) are the major group of E3s. A typical CRL complex consists of a cullin subunit (Cul1, Cul3, or Cul4), a RING protein (Hrt1/Roc1/Rbx1), an adaptor protein (Skp1 in SCF complexes, DDB1 in Cul4-based E3), and a substrate-recognition subunit such as F-box proteins (FBPs) in SCF complexes [7], BTB proteins in Cul3-type E3 complexes [8], and DCAF1s in Cul4-DDB1 E3 complexes [9–11]. In eukaryotic systems, CRLs play essential roles in many processes, including cell division, cell proliferation, cell differentiation, and circadian clock function [12]. The CRLs are activated by the neddylation process, in which Nedd8, a ubiquitin-like protein, is attached to a conserved lysine site on cullin proteins. The neddylated cullin may accelerate assembly of the CRL E3 complex, which promotes the ubiquitination of its substrate. The CSN negatively regulates the activity of CRLs by denerddylation, in which Nedd8 is cleaved from cullin proteins [4,13]. Disruption of CSN subunits generally causes hyperneddylation of Cul1 and other cullins in many organisms [4,12–16]. Genetic evidence indicates that CSN promotes CRL-mediated degradation of...
Roles of Bcl-3 in the Pathogenesis of Murine Type 1 Diabetes
Qingguo Ruan, Shi-Jun Zheng, Scott Palmer, Ruaidhri J. Carmody, and Youhai H. Chen

OBJECTIVE—It has long been recognized that autoimmunity is often associated with immunodeficiency. The mechanism underlying this paradox is not well understood. Bcl-3 (B-cell lymphoma 3) is an atypical member of the IκB (inhibitor of the nuclear factor-κB) family that is required for lymphoid organogenesis and germinal center responses. Mice deficient in Bcl-3 are immunodeficient because of the microarchitectural defects of their lymphoid organs. The goal of this study is to define the potential roles of Bcl-3 in type 1 diabetes.

RESEARCH DESIGN AND METHODS—Bcl-3–deficient NOD mice were generated by backcrossing Bcl-3–deficient C57BL/6 mice to NOD mice. Spontaneous and induced type 1 diabetes were studied in these mice by both pathologic and immunologic means. The effect of Bcl-3 on inflammatory gene transcription was evaluated in a promoter reporter assay.

RESULTS—We found that Bcl-3–deficient NOD and C57BL/6 mice were, paradoxically, more susceptible to autoimmune diabetes than wild-type mice. The increase in diabetes susceptibility was caused by Bcl-3 deficiency in hematopoietic cells but not nonhematopoietic cells. Bcl-3 deficiency did not significantly affect anti-islet Th1 or Th2 autoimmune responses, but markedly increased inflammatory chemokine and T helper 17 (Th17)-type cytokine expression. Upon transfection, Bcl-3 significantly inhibited the promoter activities of inflammatory chemokine and cytokine genes.

CONCLUSIONS—These results indicate that in addition to mediating lymphoid organogenesis, Bcl-3 prevents autoimmune diabetes by inhibiting inflammatory chemokine and cytokine gene transcription. Thus, a single Bcl-3 gene mutation leads to both autoimmunity and immunodeficiency. Diabetes 59:2549–2557, 2010

Type 1 diabetes, or insulin-dependent diabetes, is an inflammatory disease of the pancreatic islets that affects millions of people worldwide. Although the genetic and environmental factors that trigger the disease vary, the common pathologic outcome of type 1 diabetes is the destruction of insulin-producing β-cells by inflammatory cells (activated lymphoid and myeloid cells) through a process called insulitis. Development of insulitis requires coordinated expression of a large number of genes that mediate the activation, migration and effector functions of inflammatory cells (1,2). These include genes that encode cytokines, chemokines, and cytotoxic enzymes. Although it is well recognized that expression of these genes is tightly controlled at the transcriptional level, the nature of the transcription factors involved and the mechanisms of their action in type 1 diabetes are not well understood. Recent studies from several laboratories, including ours, indicate that the nuclear factor-κB (NF-κB) family of transcription factors plays crucial roles in type 1 diabetes. Thus, in both mice and humans, type 1 diabetes is associated with heightened NF-κB activation (3–6), whereas NF-κB deficiency in mice renders them resistant to the disease (7,8). Importantly, inhibiting NF-κB activities is highly effective in suppressing models of type 1 diabetes (9–11). Therefore, NF-κB has emerged as a long sought-after transcriptional regulator of type 1 diabetes.

In mammals, there are five NF-κB genes, NFKB1, NFKB2, RELA, cREL, and RELB, which in turn encode seven proteins: p105, p50, p100, p52, p65 (RelA), c-Rel, and RelB (12,13). The protein p50 is generated from limited proteasomal processing of its p105 precursor, as is p52 from its p100 precursor. The p50 and p52 proteins lack the transactivating domain found in the COOH-terminal regions of other NF-κB proteins. Therefore, their homodimers function primarily as repressors of gene transcription. The activities of NF-κB are tightly regulated by several IκB (inhibitor of κB) proteins that share high sequence and structural homologies (14). These include IκBα, IκBβ, IκBγ, IκBε, IκBζ, IκBNS, and Bcl-3. Additionally, the COOH-terminal regions of p100 and p105 also serve as IκBs. These IκB proteins contain repeated sequences of ~30 amino acids long termed ankyrin repeats, which are essential for binding to NF-κB. Unlike other IκB proteins that are located primarily in the cytoplasm, IκBζ, IκBNS, and Bcl-3 are found mainly in the nucleus and therefore are involved chiefly in regulating nuclear NF-κB activities. These nuclear IκB proteins exhibit significant differences in their NF-κB binding. Bcl-3 binds to only p50 and p52 homodimers (15,16); IκBζ binds to both p50 homodimers and p50/p52 dimers, whereas IκBNS appears to show little subunit preference (17,18). Importantly, the nuclear IκB proteins are not degraded after IKK (IκB kinase) activation, and their primary function appears to be the modulation of gene transcription in a gene- and NF-κB subunit–specific manner. Bcl-3 has been reported to be able to either activate or repress gene transcription after binding to p50 or p52 homodimers (19–22). Mice deficient in Bcl-3 are more susceptible to infectious diseases because of developmental defects of their lymphoid organs and reduced germinal center responses (23,24). These defects appear to be caused by Bcl-3 deficiency in nonhematopoietic stromal cells, and they reflect a role of Bcl-3 in activating organogenic chemokine genes.
Structural insights into a novel histone demethylase PHF8

Lin Yu1,*, Yang Wang1,*, Shuo Huang2,*, Jianjun Wang1, Zengqin Deng1, Qi Zhang1, Wei Wu1, Xingliang Zhang1, Zhao Liu1, Weimin Gong3, Zhongzhou Chen1

1State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing 100193, China; 2Department of Biochemistry and Molecular Biology, National Key Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100005, China; 3State Key Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing 100101, China

*These three authors contributed equally to this work.

Correspondence: Zhongzhou Chen
Tel: +86-10-62734078
E-mail: chenzhongzhou@cau.edu.cn
Received 12 October 2009; revised 22 November 2009; accepted 15 December 2009; published online 26 January 2010

Dynamic regulation of histone methylation/demethylation plays an important role during development. Mutations and truncations in human plant homeodomain (PHD) finger protein 8 (PHF8) are associated with X-linked mental retardation and facial anomalies, such as a long face, broad nasal tip, cleft lip/cleft palate and large hands, yet its molecular function and structural basis remain unclear. Here, we report the crystal structures of the catalytic core of PHF8 with or without α-ketoglutarate (α-KG) at high resolution. Biochemical and structural studies reveal that PHF8 is a novel histone demethylase specific for di- and mono-methylated histone H3 lysine 9 (H3K9me2/1), but not for H3K9me3. Our analyses also reveal how human PHF8 discriminates between methylation states and achieves sequence specificity for methylated H3K9. The in vitro demethylation assay also showed that the F279S mutant observed in clinical patients possesses no demethylation activity, suggesting that loss of enzymatic activity is crucial for pathogenesis of PHF8 patients. Taken together, these results will shed light on the molecular mechanism underlying PHF8-associated developmental and neurological diseases.

Keywords: PHF8 (PHD finger protein 8), histone demethylase, chromatin modification, methylated H3K9, crystal structure, X-linked mental retardation (XLMR), facial anomalies

Cell Research (2010) 20:166-173. doi: 10.1038/cr.2010.8; published online 26 January 2010

Histone methylation plays important roles in multiple biological processes including transcription, signal transduction, development, cellular proliferation and differentiation [1, 2]. For example, lysine methylation at H3K9 is usually associated with regions of transcriptionally silenced chromatin. Recent studies indicate that histone methylation can be removed by two distinct classes of histone demethylases, namely, flavin adenine dinucleotide (FAD)-dependent [3] and JMJC-containing protein families [4-7]. The reversible process of histone methylation and demethylation has been implicated in multiple biological processes including heterochromatin formation, X-inactivation, genomic imprinting and silencing of homeotic genes. Aberrant histone methylation has been linked to a number of human diseases such as leukemia and prostate cancers [8].

Human plant homeodomain (PHD) finger protein 8 (PHF8) contains a PHD-type zinc-finger domain, a JMJC domain, a vacuolar domain and a coiled coil stretch [2, 9, 10]. In vivo, the PHF8 transcript shows a ubiquitous expression pattern [11]. It is highly and widely expressed in many tissues, such as testis, lung and tumor tissues. It was found that defects in PHF8 were associated with diseases such as cleft lip, cleft palate, X-linked mental retardation (XLMR) and pathogenesis of autism [12]. These findings link PHF8 to cognitive function and midline formation [13]. It was also proposed that PHF8 participates in regulating cellular growth, development and differentiation [14]. However, as no enzymatic activities, biochemical function or molecular structure of PHF8 have been identified thus far, the function of PHF8 in development remains to be elucidated. Alignment studies and
Induced pluripotent stem (iPS) cells can be obtained by the introduction of defined factors into somatic cells. The combination of Oct4 (also known as Pou5f1), Sox2 and Klf4 (which we term OSK) constitutes the minimal requirement for generating iPS cells from mouse embryonic fibroblasts. These cells are thought to resemble embryonic stem cells (ESCs) on the basis of global gene expression analyses; however, few studies have tested the ability and efficiency of iPS cells to contribute to chimaerism, colonization of germ tissues, and most importantly, germ-line transmission and live birth from iPS cells produced by tetraploid complementation. Using genomic analyses of ESC genes that have roles in pluripotency and fusion-mediated somatic cell reprogramming, here we show that the transcription factor Tbx3 significantly improves the quality of iPS cells. iPS cells generated with OSK and Tbx3 (OSKT) are superior in both germ-cell contribution to the gonads and germ-line transmission frequency. However, global gene expression profiling could not distinguish between OSK and OSKT iPS cells. Genome-wide chromatin immunoprecipitation sequencing analysis of Tbx3-binding sites in ESCs suggests that Tbx3 regulates pluripotency-associated and reprogramming factors, in addition to sharing many common downstream regulatory targets with Oct4, Sox2, Nanog and Smad1. This study underscores the intrinsic qualitative differences between iPS cells generated by different methods, and highlights the need to rigorously characterize iPS cells beyond in vitro studies.

The pluripotency and self-renewing properties of ESCs are conferred by a set of core factors that helps determine their unique identity. Adult somatic cells can be reprogrammed to resemble ESCs when some of these key transcription factors are introduced. iPS cells can be obtained by the viral transduction of a few genes in both mouse and human cells, albeit at low efficiency. Supplementation with chemical compounds such as inhibitors to DNA methyltransferase, histone deacetylase, histone methyltransferase, mitogen-activated protein kinase (MAPK) and glycogen synthase kinase-3 (GSK3) have been reported to improve the reprogramming efficiency. Recently, iPS cells have been generated without the use of viral vectors. Although ESC-like iPS cells are routinely obtained with these methods, very few studies have carefully examined their germ-line contribution and transmission frequency. Although iPS cells have a distinct morphology and express molecular markers similar to ESCs, their ability and degree of contribution to the chimera seem highly varied. This suggests that iPS cells do not completely resemble ESCs, and there is marked disparity in the quality of iPS cell lines. Hence, we postulated that other factors in addition to the basal requirements of OSK may improve the quality of iPS cells as defined by their capacity for high germ-line competency.

We speculated that iPS-cell-reprogramming factors may share common characteristics with pluripotency-associated genes whose perturbed levels in ESCs confer resistance to differentiation. Previous studies have shown that mouse ESCs that overexpress Nanog are resistant to differentiation, express higher levels of pluripotency-associated genes, and are more effective at reprogramming somatic cells by cell fusion. Depletion of another transcription factor, Tcf3, in ESCs limits their differentiation ability and upregulates the expression of pluripotency markers such as Oct4, Sox2, Nanog and Sall4 (ref. 13). As both Nanog and Tcf3 regulate each other and are core features of the ESC transcriptional network, we speculated that similar to Nanog-overexpressing ESCs, the loss of Tcf3 may enhance fusion-mediated reprogramming of somatic hybrid cells. To test this, we used polyethylene glycol to generate drug-resistant fusion hybrids between Nanog overexpressing or Tcf3 RNA interference (RNAi) ESCs that were neomycin-resistant (Neo') and primary mouse embryonic fibroblasts (MEFs) that were puromycin-resistant (Puro') (Fig. 1a). Consistent with previous observations, Nanog-overexpressing ESCs showed enhanced reprogramming efficiency (Fig. 1b, c). Using Tcf3-deficient ESC lines, a significant increase in the number of hybrid clones was also observed (Fig. 1b, c). Karyotype analysis confirmed that these were tetraploid (Supplementary Fig. 1). The hybrids possess similar properties to the parental modified ESC lines, including their response to the lack of leukaemia inhibitory factor (LIF) and epigenetic reprogramming of the Nanog promoter (Supplementary Figs 2 and 3). We eliminated the possibility that improvements in reprogramming frequency were attributed to increased cell fusion events (Supplementary Fig. 4).

To dissect the commonalities between Nanog and Tcf3 pathways, we examined the repertoire of genes increased in Nanog overexpressing and Tcf3 RNAi ESCs that could suggest shared downstream mediators. The intersected expression profiles revealed a handful of genes such as Dazl, Fzd10, Hal, 4930502E18Rik and Erf that were upregulated in both conditions (Supplementary Tables 1 and 2). Notably, Tbx3, a transcription factor previously reported to sustain pluripotency, was strongly increased (Fig. 1d and Supplementary Fig. 5A). RNAi knockdown of Tbx3 in ESCs induced differentiation (Fig. 1e), with concomitant downregulation of pluripotency-associated genes (Supplementary Table 3). Tbx3 is also directly bound by both Nanog and Tcf3 (Fig. 1f).

In reprogrammed ESC–MEF hybrids, Tbx3 levels remained greatly increased (Supplementary Fig. 5B and Supplementary Tables 4 and 5). To test the role of Tbx3 in cell fusion-mediated reprogramming, we fused Neo' Tbx3 overexpressing ESC lines with MEFs (Supplementary Fig. 6). Indeed, there was an increase in the number of hybrids (Fig. 1b, c), which was not due to enhanced cell fusion (Supplementary Fig. 4).

1Stem Cell and Developmental Biology, Systems Biology, Genome Institute of Singapore, 138672, Singapore. 2State Key Laboratory for Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing 100193, China. 3Singapore Immunology Network, 138648, Singapore. 4Department of Biological Sciences, National University of Singapore, 117543, Singapore. 5Center for Life Sciences, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02115, USA. 6Present address: Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA.

Tbx3 improves the germ-line competency of induced pluripotent stem cells

Jianyong Han1,3, Ping Yuan1, Henry Yang4, Jinqu Zhang1, Boon Seng Soh1, Pin Li1, Siew Lam Lim1, Suying Cao1, Junliang Tay1, Yuriy L. Orlov2, Thomas Lufkin3, Huck-Hui Ng1,5, Wai-Leong Tam1† & Bing Lim1,6