Evidence for Stabilizing Selection Driving Mutational Turnover of Short Motifs in the Eukaryotic Complementary Sex Determiner (Csd) Protein

Vasco Koch,[1] Marianne Otte, and Martin Beye[1]

Institute of Evolutionary Genetics, Heinrich Heine University Duesseldorf, Universitaetsstrasse 1, 40225 Duesseldorf, Germany

ABSTRACT Short linear motifs (SLiMs) can play pivotal functional roles in proteins, such as targeting proteins to specific subcellular localizations, modulating the efficiency of translation and tagging proteins for degradation. Until recently we had little knowledge about SLiM evolution. Only a few amino acids in these motifs are functionally important, making them likely to evolve *ex nihilo* and suggesting that they can play key roles in protein evolution. Several reports now suggest that these motifs can appear and disappear while their function in the protein is preserved, a process sometimes referred to as "turnover". However, there has been a lack of specific experiments to determine whether independently evolved motifs do indeed have the same function, which would conclusively determine whether the process of turnover actually occurs. In this study, we experimentally detected evidence for such a mutational turnover process for nuclear localization signals (NLS) during the post-duplication divergence of the Complementary sex determiner (Csd) and Feminizer (Fem) proteins in the honeybee (*Apis mellifera*) lineage. Experiments on the nuclear transport activity of protein segments and those of the most recent common ancestor (MRCA) sequences revealed that three new NLS motifs evolved in the Csd protein during the post-duplication divergence while other NLS motifs were lost that existed before duplication. A screen for essential and newly evolved amino acids revealed that new motifs in the Csd protein evolved by one or two missense mutations coding for lysine. Amino acids that were predating the duplication were also essential in the acquisition of the C1 motif suggesting that the *ex nihilo* origin was constrained by preexisting amino acids in the physical proximity. Our data support a model in which stabilizing selection maintains the constancy of nuclear transport function but allowed mutational turnover of the encoding NLS motifs.

Many studies have shown that protein domains cover only a fraction of a protein's amino acid sequence, and functionally important short linear motifs (SLiMs) are often located in intrinsically unstructured regions (Björklund *et al.* 2005; Dyson and Wright 2005; Diella *et al.* 2008;

Finn *et al.* 2016). These SLiMs are usually of low complexity, comprising just a few amino acids, and play pivotal functional roles such as controlling cell-cycle progression, tagging proteins for proteasomal degradation, modulating the efficiency of translation, targeting proteins to specific sub-cellular localizations (*e.g.*, nuclear localization signals) and stabilizing scaffolding complexes (Fuxreiter *et al.* 2007; Davey *et al.* 2012; Dinkel *et al.* 2014; Davey *et al.* 2015). To date, more than 200 motif classes have been curated using experimental validation (Dinkel *et al.* 2014). Until recently, little was known about SLiM evolution, especially in comparison to global domain evolution (for a review see (Davey *et al.* 2015)). The few amino acids that are functionally important make these domains very likely to arise *de novo* (*ex nihilo*) in a protein sequence through a small number of mutations (Neduva and Russell 2005; Davey *et al.* 2012). The potential for evolutionary changes in compact and degenerate SLiMs led to the hypothesis that they play a key role in protein evolution (Neduva and Russell 2005).

Protein networks can acquire new interactions with only a few amino acid changes, thereby gaining important novel regulatory functions (Neduva and Russell 2005; Davey *et al.* 2015). There is accumulating evidence that new SLiMs can evolve *ex nihilo* (Davey *et al.* 2015). For example, several patients with the Noonan-like syndrome have independently evolved mutations in the lysine-rich repeat protein SHOC-2, which resulted in the *ex nihilo* birth of a myristoylation motif in humans (Cordeddu *et al.* 2009). Several analyses tracing the taxonomic range of motifs have shown that SLiMs are evolutionarily gained or lost in individual lineages. Extensive datasets provided by high throughput proteomics studies have shown that a large number of motifs are clade-specific (Holt *et al.* 2009; Goldman *et al.* 2014), suggesting that SLiMs have been repeatedly gained or lost. These gains and losses can be associated with functional changes of a protein. Many paralogous proteins gain distinct functionalities by gaining or losing SLiMs (Suijkerbuijk *et al.* 2012; Nguyen Ba *et al.* 2014; Di Fiore *et al.* 2015). For example, after the duplication of a Cyclin A/B ancestor, the Cyclin A regulatory subunit of the CDK protein kinase family gained an ABBA motif, allowing it to be degraded earlier than Cyclin B during prometaphase (Di Fiore *et al.* 2015). The *ex nihilo* birth of new motifs has also led to the hypothesis that motifs can appear or disappear while the protein retains its function, a process sometimes referred to as "turnover" (Moses and Landry 2010). Several reports have suggested that turnover might be a common mechanism in SLiM evolution. For example, many yeast cyclin-dependent kinase (Cdk) phosphorylation motifs are evolutionary transient, but the presence of a modification site(s) in a given protein region is conserved (Moses *et al.* 2007; Holt *et al.* 2009). However, specific experiments are needed to determine whether independently evolved motifs do indeed have the same function, to conclusively determine whether the process of turnover actually occurs (Moses and Landry 2010; Davey *et al.* 2015).

In this study, we experimentally found evidence for such a turnover process for nuclear localization signal (NLS) motifs during the postduplication divergence of the Complementary sex determiner (Csd) and Feminizer (Fem) proteins in the honeybee lineage (*Apis mellifera*). We tested the amino acid sequences of a possible most recent common ancestor (MRCA) and demonstrated the gain of new NLS motifs by investigating the required amino acid changes.

Csd and Fem proteins are SR-type splice regulators that control sex determination via alternative splicing of the *fem* and *doublesex* (*dsx*) transcripts in the honeybee (*Apis mellifera*) (Beye *et al.* 2003; Gempe *et al.* 2009; Beye *et al.* 2013). The paralogous genes *csd* and *fem* evolved recently in the honeybee lineage by gene duplication (Beye *et al.* 2003; Hasselmann *et al.* 2008a; Hasselmann *et al.* 2010; Koch *et al.* 2014) (Figure 1), resulting in *csd* evolving as a new primary signal of sex determination in the honeybee (*Apis mellifera*). Meanwhile, the Transformer (Tra) proteins, which are orthologs of Fem, have retained their roles as splicing regulators (Hoshijima *et al.* 1991; Li and Bingham 1991; Pane *et al.* 2002; Hediger *et al.* 2010; Verhulst *et al.* 2010). Splice regulators need to be transported from the cytosol into the nucleus to perform the splicing process (though antibody staining of the location is lacking due to the absence of a specific antibody). This nuclear transport is controlled by NLS motifs that can vary in their amino acid composition (see (Fried and Kutay 2003; Yasuhara *et al.* 2009) for a review). Typical NLS motifs are dominated by basic amino acids that bind to importins, protein complexes that support the direct transport of proteins from the cell plasma to the nucleus through the nuclear pore (Fagerlund *et al.* 2002; Marfori *et al.* 2011). For example, bipartite NLS motifs consist of two clusters, each consisting of two to four basic amino acids (either lysine or arginine) separated by 10 amino acids (Dingwall and Laskey 1991; Robbins *et al.* 1991). A protein can carry several NLSs
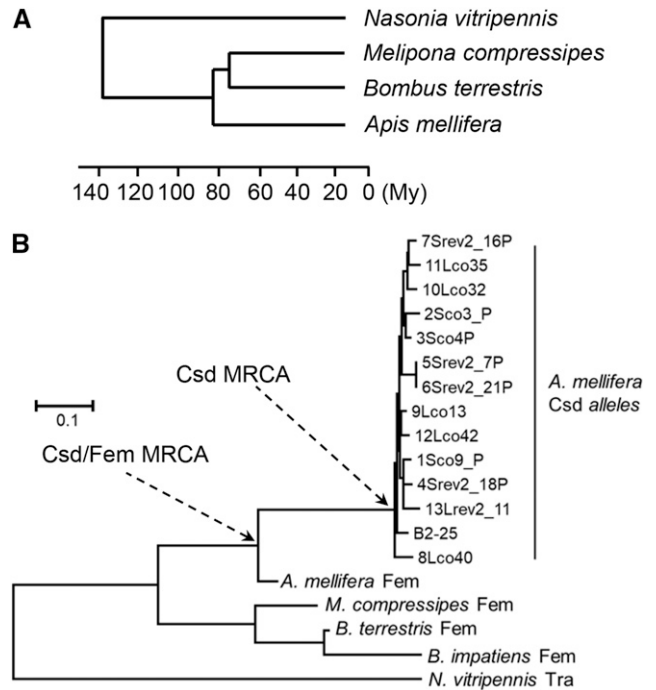

Figure 1 The evolutionary relationship of the bumble bee and honeybee species (*Apis mellifera*) and the post duplication divergence of the Fem and Csd proteins in the honeybee lineage (Hasselmann *et al.* 2008a). A: Time calibrated phylogenetic tree (in millions of years, My) of three bee species and one wasp species (Ramírez *et al.* 2010; Cardinal and Danforth 2013; Romiguier *et al.* 2016). B: The evolutionary history of protein sequences of selected bee and wasp species with the divergence of the Fem proteins and Csd alleles in the honeybee (*Apis mellifera*) lineage. The tree was inferred using the Neighbor-Joining method. Evolutionary distances were computed using the JTT matrix-based method and the modeling of rate variation among sites with a gamma distribution. The units are the number of amino acid substitutions per site. The species were: *Apis mellifera*, *Bombus terrestris*, *Bombus impatiens*, *Melipona compressipes*, *Nasonia vitripennis*. The Most Recent Common Ancestor (MRCA) of Csd and Fem proteins and Csd alleles in the honeybee lineage are marked with an arrow.

(Walther *et al.* 2005; Buck *et al.* 2006) that can be functional redundant (Südbeck and Scherer 1997; Parker *et al.* 2000). While individual NLSs are sufficient to promote nuclear transport of the protein, they can functionally be replaced by other NLSs present elsewhere in the protein.

MATERIALS AND METHODS

Cloning of nucleotide sequences
We first introduced Myc, Rubia and EGFP sequences into the PIZ/V5-His vector (Invitrogen, Carlsbad, CA, USA). The PIZ/V5-His vector was digested with *NotI* and *XbaI* to add a multiple cloning site, which was generated by polymerase chain reaction (PCR) using the oligonucleotide primers #27/#28 (Table S2). The resulting vector was digested using *XbaI* and *SacII*, and then the ORF of enhanced green fluorescent protein (EGFP) was amplified using oligonucleotide primers #01/#02 (Table S2) and inserted into the vector (Cormack *et al.* 1996). Subsequently, the ORF of the Rubia fluorescent protein (Schulte *et al.* 2013) was ligated into the pIZ/V5-His Spacer-EGFP plasmid using the restriction sites *EcoRI* and *NotI*. We also inserted a Myc tag and an

AflII site using amplicons from oligonucleotide primers #37/#38 (Table S2), so the encoded proteins were fused with the N-terminus of the Rubia protein. The resulting plasmid (pIZ/V5-His Myc-*AflII*-Rubia-EGFP and Fig. S1) was used as a vector for the different *csd* and *fem* and derived sequences, as shown in Figures 2, 3, S2, S3, and S4, via *AflII* and *EcoRI* restriction sites. We also amplified the open reading frame (ORF) of Histone H2B from *Arabidopsis thaliana* using oligonucleotide primers #026/#027 (Table S2) and then inserted it into the pIZ/V5-His-Spacer-Cerulean plasmid via the *EcoRI* and *NotI* restriction sites. The pIZ/V5-His-Spacer-Cerulean plasmid was generated by cloning the ORF of the Cerulean fluorescent protein (Rizzo *et al.* 2004). For the analyses of the full-length Csd protein, we generated five mutational variants of the NLS sequences of *csd* allele B2–25. Briefly, we inserted the NLS sequences using the *AflII* and *EcoRI* restriction sites. The sequences with nucleotide changes were generated by PCRs with no template. The pIZ/V5-His-Csd MRCA of *csd* NLS 1-2-Rubia vector was generated in 3 steps. We amplified the *csd* B2–25 allele (#1a/#1d, Table S2) and inserted the amplicons into the pIZ/V5-His Myc-*AflII*-Rubia-EGFP vector via its *AflII* and *EcoRI* restriction sites. Next, via the *SapI* and *EcoRI* restriction sites, we inserted into this vector PCR amplicons generated with no template using oligonucleotide primers #2a/#2b (Table S2). In the last step, we introduced the amplicon of the *csd* B2–25 allele (#3a/#1c, Table S2) via *BbsI* and *EcoRI* restriction sites. The pIZ/V5-His-Csd NLS 1–3 mutated-Rubia vector was generated in 4 steps. We amplified the *csd* B2–25 allele (#1b/#1d, Table S2) and introduced the amplicon via the *AflII* and *EcoRI* restriction sites into the pIZ/V5-His Myc-*AflII*-Rubia-EGFP vector. In the same vector, we inserted the amplicons of oligonucleotide primers #2a/#2d (Table S2), generated with no template, via the *SapI* and *EcoRI* restriction sites. In the next step, we inserted a second amplicon (#3b/#3c, Table S2) generated with no template via the vector's *BBSI* and *XhoI* restriction sites. In the last step, we inserted the amplicon of the *csd* B2–25 allele (#3d/#1c, Table S2) via *XhoI* and *EcoRI* restriction sites. To generate the pIZ/V5-His-Csd NLS 1 and 2 mutated-Rubia vector, we used the pIZ/V5-His vector, into which the amplicon of oligonucleotides #1b/#1d (Table S2) was already inserted. First, we inserted amplicons (oligonucleotide primers #2a/#2d and no template, Table S2) via the *SapI* and *EcoRI* restriction sites. Second, we inserted amplicons of the *csd* B2–25 allele (#3a/#1c, Table S2) via the *BbsI* and *EcoRI* restriction sites. We also generated pIZ/V5-His-Csd NLS 1 and 3 mutated-Rubia vectors. We used the pIZ/V5-His vector (which already possessed the amplicon from oligonucleotide primers #1b/#1d, Table S2). We inserted two PCR products that were generated without a template: the amplicon of #2a/#2b (Table S2) via the *SapI* and *EcoRI* restriction sites and the amplicon of #3b/#3c (Table S2) via the *BBSI* and *XhoI* restriction sites. In the last step, we inserted into the *XhoI* and *EcoRI* restriction sites the amplicon of the *csd* B2-25 allele generated with oligonucleotides #3d/#1c (Table S2). To construct the pIZ/V5-His-Csd NLS 2 and 3 mutated-Rubia vector we used a pIZ/V5-His vector that already possessed the amplicon of oligonucleotides #1a/#1d (Table S2). We introduced two amplicons generated with no template: the amplicon of #2a/#2d (Table S2) via the *SapI* and *EcoRI* restriction sites and the amplicon of #3b/#3c (Table S2) via the *BbsI* and *XhoI* restriction sites. Next, we inserted the amplicon of the *csd* B2-25 allele generated with #3d/#1c (Table S2) via the *XhoI* and *EcoRI* restriction sites. Since the expression of the Csd full-length protein from plasmids was very low and difficult to detect in *Sf*21 cells, we expressed the full-length Csd proteins using the baculovirus expression system (Invitrogen, Carlsbad, CA, USA). We cloned the *csd* sequences described above into pFastBac HTa vectors (Invitrogen, Carlsbad, CA, USA) using the restriction sites *MfeI* and

SalI. Finally, we inserted each *csd* sequence in-frame with the Rubia fluorescent protein ORF (Fig. S1).

Cell culture, transfection and microscopic analysis

*Sf*21 cells were adherently cultured in 10 ml of Spodopan medium (PAN-Biotech, Aidenbach, Germany) containing 10 μg/ml of gentamycin (Carl Roth, Karlsruhe, Germany) in 250 ml cell culture flasks (75 cm²; Greiner Bio-One). We transfected 1×10^6 *Sf*21 cells with 2.5 μg of plasmid DNA via Roti-Insectofect reagent following the manufacturer's instructions (Carl Roth, Karlsruhe, Germany). Baculovirus stocks were incubated following the Bac-to-Bac Baculovirus Expression System manual from Invitrogen (Invitrogen, Carlsbad, CA, USA).

We examined transfected *Sf*21 cells 24 to 72 hr after transfection using a confocal fluorescence microscope (Zeiss LSM510META, Carl Zeiss Microscopy, Jena, Germany) and an Achroplan 40×/0.8 W objective. Fluorescence was detected at wavelengths of 561 nm (Rubia) and 458 nm (Cerulean).

Sequence analysis

We determined ancestral sequences (Fig. S7) using the ANC-GENE software (Zhang and Nei 1997) and the MEGA5 software package (Tamura *et al.* 2011) using nucleotide sequence alignments. We used the distance-based Bayesian method to infer the ancestral amino acid states and then inferred the underlying nucleotide sequences by fixing the inferred amino acids. ANC-GENE and MEGA5 yielded identical results for the ancestral amino acid states of the NLS elements. The following *csd*, *fem* and *tra* sequences were used for the analysis (GenBank accession numbers): *Apis mellifera csd* alleles: B2-25 (AY569703.1); Sco9_P (EU100895.1); Sco3_P (EU100893.1); Sco4P (EU100894.1); Srev2_18P (EU100898.1); Srev2_7P (EU100896.1); Srev2_21P (EU100896.1); Srev2_16P (EU100897.1); Lco40, (EU100888.1); Lco13 (EU100885.1); Lco32 (EU100886.1); Lco35 (EU100887.1); Lco42 (EU100889.1); Lrev2_11 (EU100890.1). *Apis mellifera fem* (AY569719.1); *Bombus terrestris fem*: (100628566); *Bombus impatiens* (100742483) *Nasonia vitripennis transformer* (*tra*)(EU780924.1); *Melipona compressipes*; *fem* (EU139305.1). To test whether the similarity between the sequences was sufficiently high to generate informative MRCA nucleotide and amino acid sequences (Table S1), we performed tests on the saturation of substitutions at the 3rd codon position between *Bombus* and *Apis* sequences using the methods implemented in the DAMBE6 program (Xia *et al.* 2003; Xia 2017) and examined the number of synonymous substitutions per synonymous site between nucleotide sequences (*ds*) using the Nei-Gojobori model, which is implemented in the MEGA5 software package (Tamura *et al.* 2011). We used NLS mappers to identify possible NLSs in the Csd amino acid sequence (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi (Kosugi *et al.* 2009); http://www.moseslab.csb.utoronto.ca/NLStradamus/ (Nguyen Ba *et al.* 2009), last successful access October 2017). and IUPred2A program to identify disordered regions (Mészáros *et al.* 2018) (https://iupred2a.elte.hu/ last successful access August 2018). The evolutionary history of the Fem and Csd protein sequences were inferred with models implemented in the MEGA5 software package (Tamura *et al.* 2011). We used Neighbor-Joining method with the tree drawn to scale, with branch lengths in the same units as those of the evolutionary distances. Maximum likelihood fits were run to identify best fit amino acid substitution models. Evolutionary distances were computed using the JTT matrix-based method and the modeling of rate variation among sites with a gamma distribution (shape parameter = 1). All positions containing gaps and missing data were eliminated.
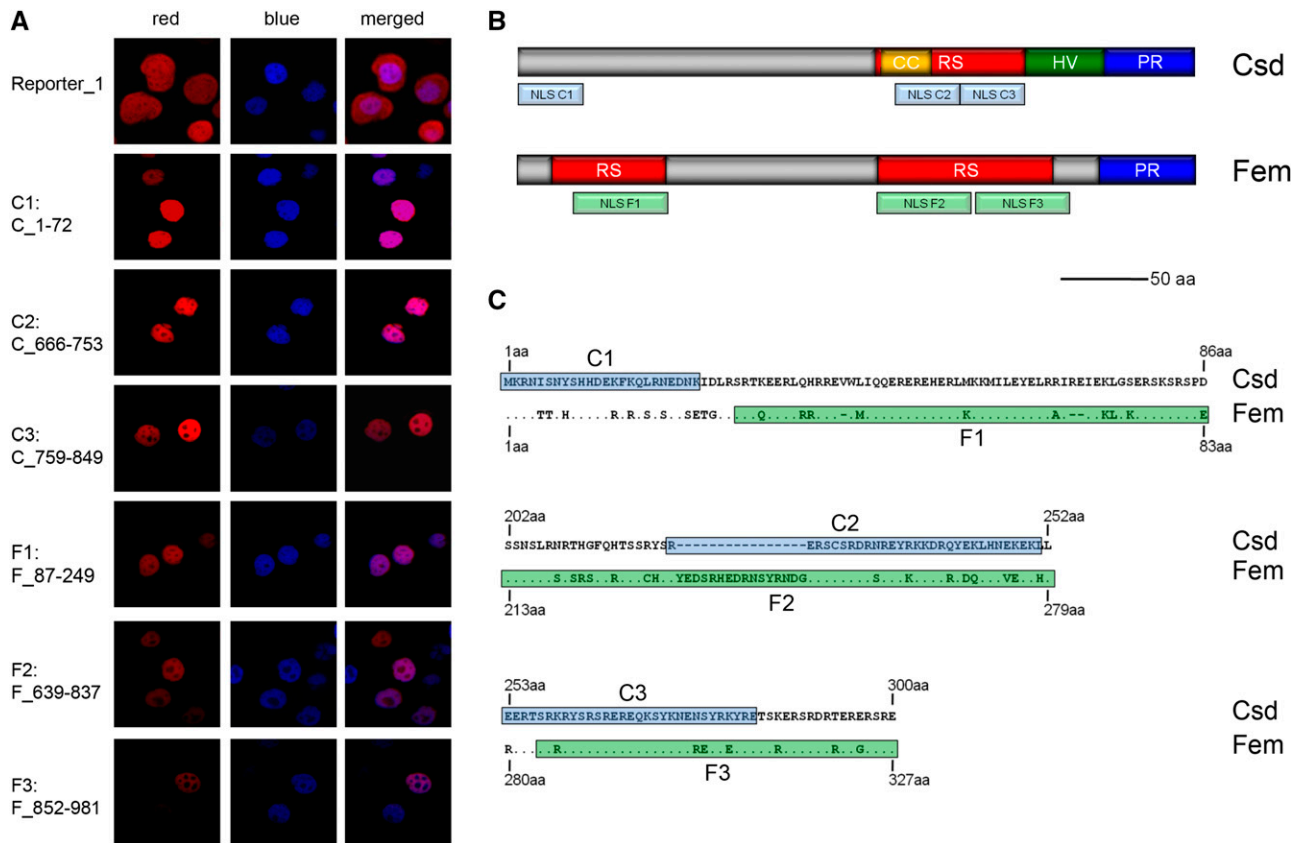
Figure 2 Minimal segments of the Csd and Fem proteins providing nuclear transport activity. A: The identified minimal segments of the Csd and Fem proteins that are sufficient to localize the reporter protein Rubia into the nucleus of *Sf*21 insect cells (red fluorescence signal in the pictures). The positive marker of nuclear localization (histone H2B fused with the Cerulean protein) was detected via its blue fluorescence signal. The nucleotide position encoding the tested segment of the allele *csd* B2–25 is shown to the left. The reporter protein without Csd- or Fem-derived segments is shown at the top (Reporter_1). Other segments examined are shown in Figs. S2 and S3. The different fusion proteins used are schematically presented in Fig. S1. B: The location of the segments providing nuclear localization function to the Csd and Fem proteins. The arginine-serine-rich region (RS domain), the coiled-coil motif (CC), the hypervariable region (HV region, confined to the Csd protein), and the proline-rich region (PR region) are shown. C: Amino acid sequence alignments of the minimal Csd and Fem segments. The C1 to C3 segments of the Csd protein are marked in light blue, while the F1 to F3 segments of the Fem protein are colored in light green.

Data Availability Statement

Plasmids are available upon request. The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures and supplementary data set available at Figshare: https://figshare.com/s/944fea8a9dfa2045b1f8.

RESULTS

New NLS motifs evolved in the Csd protein

To investigate the evolution of their localization function post divergence, we characterized minimal segments of the Fem and Csd proteins along with MRCA sequences and tested whether they were sufficient to direct localization into the nucleus. The sequences in this study were fused to the red fluorescent protein Rubia with spacer sequences, and then they were expressed in *Sf*21 insect cells using expression or baculovirus systems (Figure 2). We tested three minimal segments for the Csd protein (C1: aa 1–24, 24 aa long; C2: aa 222–251, 30 aa long; C3: aa 253–283, 31 aa long) and three for the Fem protein (F1: aa 29–83, 55 aa; F2: aa 213–279, 67 aa; F3: aa 284–327, 44 aa) that were sufficient for nuclear localization (Figures 2, S2 and S3). We used the Histone H2B protein from *Arabidopsis thaliana* as a marker of nuclear localization because histone proteins are typical proteins that need to be transported

into the nucleus. Each of these constructs co-localized with the Histone H2B protein from *Arabidopsis thaliana*, fused with the blue fluorescent protein Cerulean, which we used as a positive marker for nuclear transport (Figures 2 and S1). As a control, the reporter protein was expressed alone (Figures 2 and S1), and it was not transported into the nucleus. When we further reduced the size of the Csd and Fem segments, the proteins were not transported to the nucleus but were instead detected in the cytoplasm (Figs. S2 and S3). These segments are all located in disordered domains of the proteins consistent with the location of short motifs in other proteins (Björklund *et al.* 2005; Dyson and Wright 2005; Diella *et al.* 2008; Finn *et al.* 2016). Next, we inferred the sequence of the Csd and Fem MRCA (Csd/Fem MRCA, Figure 1B) and that of Csd (Csd MRCA) by applying a Bayesian approach to alignments of the coding sequences comprising multiple *csd* alleles and *fem* sequences from the honeybee and *fem* sequences from three other bee species and the ortholog *tra* sequence from the wasp *Nasonia* (Figure 1; Table S1). MRCA sequences of the Csd/Fem C1, C2 and C3 segments, representing the ancestral state prior to the duplication and divergence of Csd and Fem proteins, were not transported to the nucleus (Csd/Fem MRCA_C1, _C2 and _C3; Figure 3 A, B and C). The C1, C2 and C3 homologous segments of the Fem protein (Fem_C_1–72, Fem_C_666–753 and
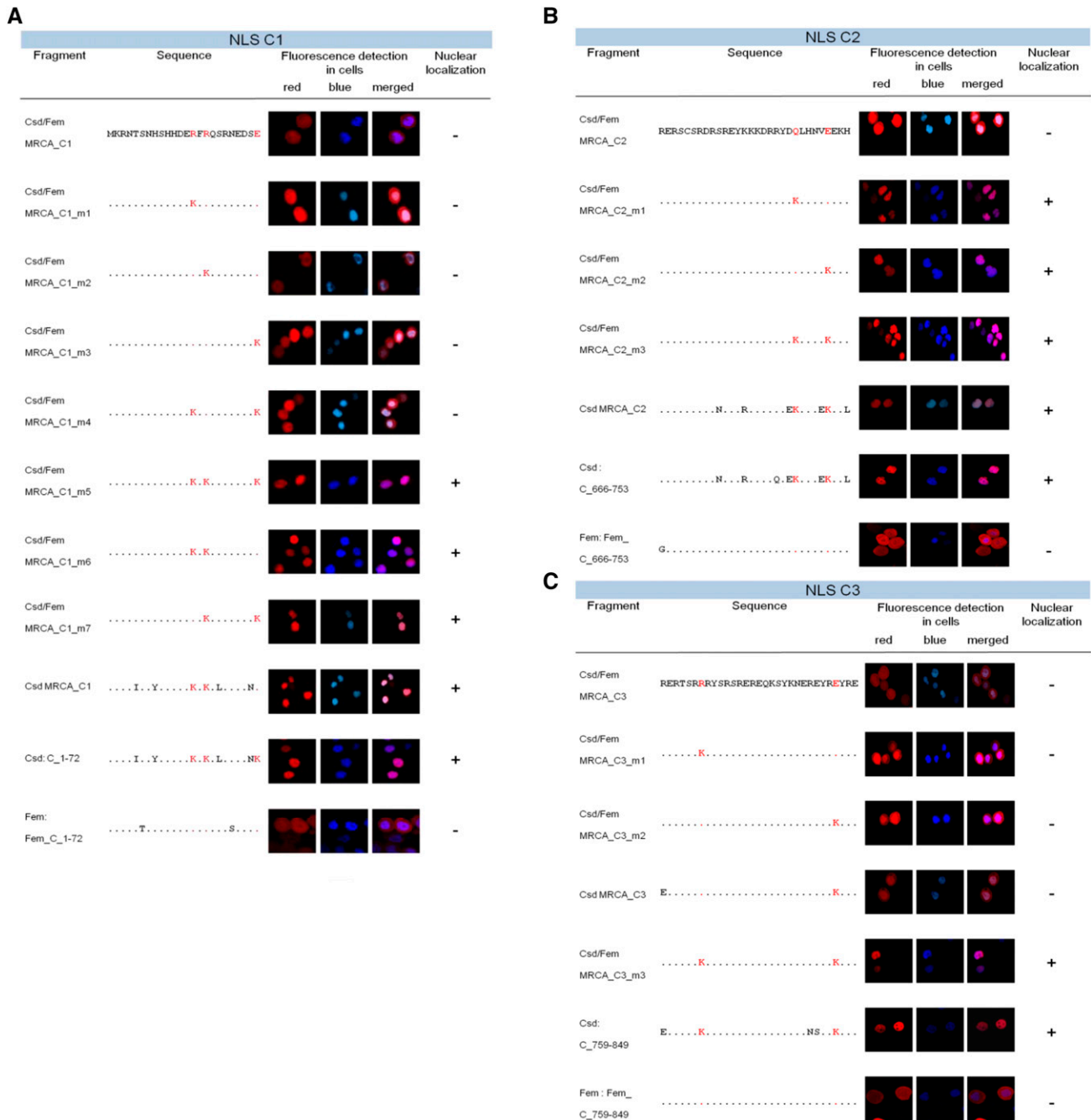
A

NLS C1

Fragment	Sequence	Fluorescence detection in cells			Nuclear localization
		red	blue	merged	
Csd/Fem MRCA_C1	MKRNTSNHSHHDERFRQSRNEDSE				−
Csd/Fem MRCA_C1_m1	..............K...				−
Csd/Fem MRCA_C1_m2	..............K...				−
Csd/Fem MRCA_C1_m3	...................K				−
Csd/Fem MRCA_C1_m4	..........K.......K				−
Csd/Fem MRCA_C1_m5	..........K.K.....K				+
Csd/Fem MRCA_C1_m6	..........K.K.....				+
Csd/Fem MRCA_C1_m7	...............K...K				+
Csd MRCA_C1	....I..Y....K.K.L....N.				+
Csd: C_1-72	....I..Y....K.K.L....NK				+
Fem: Fem_C_1-72	.....T..........S....				−

B

NLS C2

Fragment	Sequence	Fluorescence detection in cells			Nuclear localization
		red	blue	merged	
Csd/Fem MRCA_C2	RERSCSRDRSREYKKKDRRYDQLHNVEEKH				−
Csd/Fem MRCA_C2_m1	.....................K......				+
Csd/Fem MRCA_C2_m2	.....................K...				+
Csd/Fem MRCA_C2_m3	...............K...K...				+
Csd MRCA_C2	.........N...R.....EK...EK..L				+
Csd: C_666-753	.........N...R...Q.EK...EK..L				+
Fem: Fem_C_666-753	G............................				−

C

NLS C3

Fragment	Sequence	Fluorescence detection in cells			Nuclear localization
		red	blue	merged	
Csd/Fem MRCA_C3	RERTSRRRYSRSREREQKSYKNEREYREYRE				−
Csd/Fem MRCA_C3_m1	......K.....................				−
Csd/Fem MRCA_C3_m2	......................K...				−
Csd MRCA_C3	E.....................K...				−
Csd/Fem MRCA_C3_m3	......K...............K...				+
Csd: C_759-849	E.....K...........NS..K...				+
Fem: Fem_C_759-849	....				−

Figure 3 Evolutionary origin of NLS motifs in the C1, C2 and C3 segments through amino acid substitutions. A: Essential amino acid substitutions in the C1 segment for gaining nuclear localization activity through post-duplication divergence. B: Essential amino acid substitutions in the C2 segment for gaining nuclear localization activity through post-duplication divergence. C: Essential amino acid substitutions in the C3 segment for gaining nuclear localization activity through post-duplication divergence. The sites and the evolved amino acids in the Csd protein involved in the gain of nuclear transport function are shown in red letters. The sites are amino acid positions 14, 16 and 24 for the C1 segment; 243 and 248 for the C3 segment; and 259 and 280 for the C3 segment, with allele B2–25 as a reference. The red fluorescence signal in the pictures of Sf21 cells indicates the C-segment variants fused to the Rubia protein. The blue signal indicates the nuclear-localized H2B_Cerulean protein.

Fem_C_759–849) were also not sufficient to direct nuclear localization (Figure 3). These results together indicate that three NLS motifs evolved during the post-duplication divergence of the Csd protein. Further, we examined whether this divergence occurred in the branch before or in the branches after the divergence into multiple Csd alleles (Figure 1B). This examination is possible because Csd alleles have been maintained by strong balancing selection over millions of years

(Hasselmann *et al.* 2008b; Lechner *et al.* 2014). To interrogate this, we examined the nuclear transport of the MRCA sequences of the Csd alleles (Csd MRCA, Figure 1B) in the C1, C2 and C3 segments. We observed that Csd MRCA_C1, Csd MRCA_C2 and Csd MRCA_C3 were transported into nucleus (Figure 3). Together, these results indicate that C1, C2 and C3 NLS motifs evolved after duplication, but prior to the divergence of different Csd alleles. As for the F elements

■ **Table 1 The amino acid (aa) and nucleotide states at the sites of the functionally relevant lysines before the Csd/Fem duplication and divergence event (Csd/Fem MRCA) and before the Csd allele divergence (Csd MRCA)**

		aa/codon of Csd B2–25	aa/codon of Csd/Fem MRCA ($P > 0.9$)	aa/codon of Csd MRCA ($P > 0.9$)	aa/codon polymorphisms (frequency %)[3]
NLS	C1	K^{14}	R	K	K (100%)
		AAA	AGA	AAA	AAA
		K^{16}	R	K	K (100%)
		AAA	AGA	AAA	AAA
		K^{24}	E	E	E (57%) K (43%)
		AAA	GAA	GAA	GAA AAA
	C2	K^{243}	Q	K	K (100%)
		AAA	CAA	AAA	AAA
		K^{248}	E	K	E (7%) K (93%)
		AAA	GAA	AAA	GAA AAA
	C3	K^{259}	R	K	E (14%) K (79%) N (7%) GAG AAG AAC
		AAG	AGG	AAG[2]	
		K^{280}	E [1]	K	K (21%) K (79%)
		AAG	GAA	AAA	AAG AAA

[1] Ambiguous codon ($P < 0.9$) due to indels that occurred with outgroup sequence comparison.
[2] The predicted codon was R (AGG), $P < 0.6$ using ANC-GENE (Zhang and Nei 1997), and K (AAG), $P > 0.9$ using MEGA (Tamura *et al.* 2011). From the more parsimonious number of mutations required to produce the other polymorphism (GAG and AAC), we suggest that the aa/codon of the *csd* MRCA is K/AAG.
[3] Estimated from a random sample of 14 *csd* alleles.

of the Fem protein, we were only able to infer the Csd/Fem MRCA sequence of the F1 segment. The Csd/Fem MRCA_F1 sequence was transported into the nucleus (Fig. S4), indicating that this transport function is conserved and pre-dated the duplication and divergence of the Csd and Fem proteins. Independent deletions and insertions of the homologous F2 and F3 segments in different bee lineages made it impossible to infer the F2 and F3 Csd/Fem MRCA sequences. To further explore the history of the F2 and F3 segments, we examined the transport function of the homologous F2 and F3 segments of the bumblebee *Bombus terrestris fem* gene. The F3 segment of *B. terrestris* was transported into the nucleus, while the F2 element was not (Fig. S4), suggesting that the transport function of the F3 segment also pre-dated the duplication event.

Nuclear transport functions repeatedly evolved via one or two missense mutations coding for lysine

Next, we screened for amino acids that evolved after duplication in the Csd protein resulting in the gain of nuclear transport. We introduced amino acids into the Csd/Fem MRCA sequence that evolved during the post-duplication divergence, but prior to the divergence of the Csd alleles (*i.e.*, changes detected between the Csd/Fem MRCA and Csd MRCA sequences) and tested for nuclear transport. To understand the evolution of the elements also after allele divergence we introduced those amino acids that evolved in lineage of the allele CsdB2-25 (Figures 3, S5).

For the C1 segment, we observed that two post-duplication changes (replacement of R^{14} (arginine) with K^{14} (lysine) and R^{16} with K^{16}) were sufficient to direct nuclear transport (Figure 3A). The E^{24} (glutamic acid) to K^{24} change that evolved after the divergence of Csd alleles was also sufficient to mediate nuclear transport, but only together with the R^{16} to K^{16} replacement. Nucleotide sequence analysis (Table 1) revealed that each of the amino acid changes resulted from single nonsynonymous nucleotide replacements. Together, these results suggest that a new NLS motif in C1 evolved during the post-duplication divergence of the Csd and Fem genes by two missense mutations coding for the amino acid lysine.

For the C2 segment, we discovered that replacing either Q^{243} (glutamine) with K^{243} or E^{248} with K^{248} in the Csd/Fem MRCA_C2 sequence was sufficient to mediate nuclear localization (Figure 3B).

Both lysines evolved by single nucleotide changes (Table 1). K^{243} evolved post duplication while K^{248} either evolved post duplication or post divergence of the Csd alleles (Table 1). These results suggest that a new NLS motif in the C2 segment evolved by a single missense mutation coding for lysine.

For the C3 segment, we found that substituting R^{259} with K^{259} and E^{280} with K^{280} in the Csd/Fem MRCA_C3 sequence was sufficient to direct nuclear localization (Figure 3C). Both lysine-encoding codons evolved by single nucleotide changes (Table 1). K^{259} and K^{280} (Table 1) evolved post duplication. Together, these results indicate that the new NLS motif in the C3 segment evolved by two single missense mutations coding for lysine.

The evolution of the NLS motif in C1 was constrained by preexisting amino acids

Our results indicate that new NLS motifs can evolve via one or two missense mutations. We next investigated whether there are constraints on the pre-existing sequence on where in a protein the new motifs can evolve. To address this question, we used a C1 Csd/Fem MRCA sequence with the lysine changes K^{14} and K^{16}. This sequence represented the ancestral sequence (prior to duplication) with the two mutations that gained the new nuclear transport function (Fig. S7). We then replaced amino acids that pre-existed prior to duplication and tested whether these ancestral amino acids were essential for the gain of nuclear transport together with the newly evolved lysine. Replacing basic amino acids lysine (K) or arginine (R) at sites 2, 3 and 19 with either of the neutral amino acids alanine or glycine resulted in loss of nuclear transport function (Fig. S7). Substituting glutamic acid (E) at site 13 with alanine or glycine had no such effect (Fig. S7). These observations indicate that several basic amino acids existed prior to duplication that became essential for the rise of the new NLS motif. This result indicates that the origin of a new motif by two mutations also required other basic amino acids in close proximity that pre-existed in the ancestral sequence.

Newly evolved lysines were essential in producing three NLS motifs in the Csd protein

We next investigated whether the newly evolved NLS motifs in the C1, C2 and C3 segments were indeed active in the full-length Csd protein.
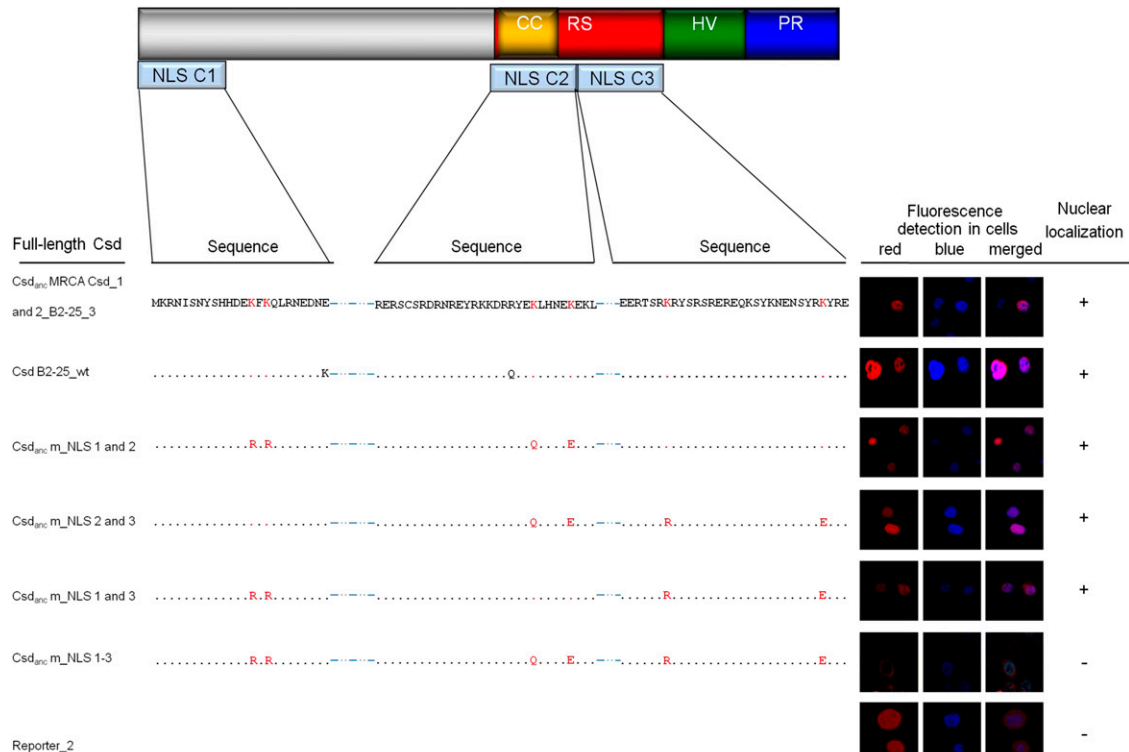
Figure 4 Essential role of the newly evolved lysines in the transport of the Csd protein. A: Schematic view of the domain structure of the Csd protein with the relative positions of the NLS elements. The arginine-serine-rich region (RS domain), the coiled-coil motif (CC), the hypervariable region (HV region), and the proline-rich region (PR region) are shown. B: Sequence changes in the C1, C2 and C3 segments of the Csd protein and their consequences on nuclear transport. The newly evolved lysines that gave rise to the NLS function, as determined in the segment analysis, were reverted to the ancestral state (Csd/Fem MRCA sequence) in each C segment. The sites that were changed are colored in red (codon sites 14, 16 for C1; 243 and 248 for C2; and 259 and 280 for C3). The red fluorescence signal in the pictures of *Sf*21 cells indicates the Csd protein variants fused to the Rubia protein. The blue signal represents nuclear H2B_Cerulean protein marking the nucleus. The images behind each sequence show the localization of that protein when expressed in *Sf*21 cells. The proteins (Fig. S1D) were fused with the Rubia protein and detected via red fluorescence. The nucleus was detected as the blue fluorescence of the histone H2B-Cerulean protein. Reporter_2 is the reporter protein not fused with Csd (fig. S1E).

To investigate this, we mutated the newly evolved lysines in each motif and examined the nuclear transport of the entire protein.

First, we generated and expressed the Csd $_{MRCA\ Csd\_1\ and\ 2}$ sequences from Csd allele B2–25. This protein contained the amino acid sequences at the C1, C2 and C3 segments from before the Csd allele divergence. We observed that the Csd $_{MRCA\ Csd\_1\ and\ 2}$ proteins were transported into the nucleus (Figure 4), suggesting that the transport function was present prior to the Csd allele divergence. Next, we expressed the Csd $_{m\_NLS1-3}$ sequence, in which all six functionally relevant lysines of all three segments (K^{14} and K^{16} for C1; K^{243} and K^{248} for C2; and K^{259} and K^{280} for C3) were reverted to their ancestral state prior to duplication. This Csd $_{m\_NLS1-3}$ sequence was not transported to the nucleus (Figure 4) suggesting that no other NLS motifs are present in the Csd protein. Next, we reintroduced, stepwise for each motif, the evolved lysines into Csd $_{m\_NLS1-3}$ and tested for gain of transport function. Introducing lysines (K^{14} and K^{16}) of C1 into the Csd $_{m\_NLS1-3}$ protein resulted in nuclear transport (Csd $_{m\_NLS\ 2\ and\ 3}$; Figure 4) indicating that the evolution of those two lysines was sufficient to mediate nuclear transport of the Csd protein. Introducing either lysines K^{243} and K^{248} of C2 or lysine K^{259} and K^{280} of C3 into the Csd $_{m\_NLS1-3}$ protein also mediated nuclear transport (Csd $_{m\_NLS\ 1\ and\ 3}$, Csd $_{m\_NLS1\ and\ 2}$; Figure 4), suggesting that the evolved lysines in the C2 or C3 motifs are each sufficient to mediate the nuclear transport of the Csd protein. Together our results

indicate that each of the three newly evolved motifs can mediate nuclear transport of the Csd protein.

DISCUSSION

Our results provide experimental evidence for the mutational turnover of NLS motifs after the divergence of the Csd and Fem proteins. Three NLS motifs newly appeared in the Csd proteins by one or two point mutations, while the evolutionarily older motifs that existed in the common ancestor of the Csd/Fem proteins were lost, demonstrating motif turnover and the preservation of nuclear transport. Several reports have suggested that turnover might be a common mechanism in SLiM evolution (Moses and Landry 2010; Davey *et al.* 2015). For example, studies on yeast cyclin-dependent kinase (Cdk) have identified specific phosphorylation motifs that have changed, while the presence of modification sites in given protein regions has been conserved (Moses *et al.* 2007; Holt *et al.* 2009). However, with data presented here we have demonstrated that newly evolved motifs can indeed have the same function as their ancestral sequence, which provides experimental support for the turnover model.

Our data support a model in which stabilizing selection maintains the constancy of nuclear transport function but allowed mutational turnover of the encoding NLS motifs. One driving force for this turnover of motifs is the ease by which new motifs can evolve *ex nihilo* through a

small number of mutations (Lynch 2007). Random genetic drift or further adaptive adjustment and selection for new functions are possible evolutionary forces that may drive new variants in the population to fixation. Interestingly, turnover due to stabilizing selection at the level of gene regulation is a common model for the evolution of transcription factor binding sites *in cis*-regulatory modules. There is strong support from the results of genome-wide and single gene-based studies (Ludwig *et al.* 2000; Dermitzakis and Clark 2002; Moses *et al.* 2006; Bradley *et al.* 2010; Arnold *et al.* 2014). For example, despite high sequence divergence, the *eve* stripe enhancer regions from closely related species drive nearly indistinguishable expression patterns in *Drosophila melanogaster* (Ludwig *et al.* 1998), while the specific transcription factor binding sites responsible for their expression patterns seem to have changed during evolution (Ludwig *et al.* 2000). Our results on NLS motifs suggest a related phenomenon of stabilizing selection for the evolutionary turnover of protein SLiMs.

Our results on the mutational steps essential for the creation of three NLS motifs further support the model of *ex nihilo* SLiM evolution by a small number of mutations (Neduva and Russell 2005; Davey *et al.* 2015). Only two replacements in the C1 and C3 and one replacement in C2 segment, all with the amino acid lysine, were sufficient to give birth to new NLS motifs in the Csd protein. These changes required only single nonsynonymous mutations, suggesting that new motifs may indeed arise by chance (Davey *et al.* 2015). Further, for the C1 and C3 motifs a single mutation alone was not sufficient to direct even slight nuclear transportation, indicating that partial gain of function was not driving motif acquisition. We also revealed that amino acids that were predating the duplication were essential in the acquisition of the C1 motif suggesting that the *ex nihilo* origin of SLiMs is constrained by preexisting amino acids in the physical proximity.

The three newly evolved NLS motifs were functionally redundant in our transport assay, suggesting that all of them are functionally relevant in honeybees. This finding is consistent with reports of other protein families having multiple NLSs (Walther *et al.* 2005; Buck *et al.* 2006); other members of the protein family from the present study are all splice regulators that are transported into the nucleus (Hoshijima *et al.* 1991; Li and Bingham 1991; Pane *et al.* 2002; Hediger *et al.* 2010; Verhulst *et al.* 2010). Neither our algorithmic nor our data bank-based motif analyses have predicted the identified NLS motifs. However, the pattern of the essential amino acid lysine to the left and right of the minimal segments in all three motifs suggest that the newly evolved motifs belong to the class of bipartite NLSs (Dingwall and Laskey 1991; Makkerh *et al.* 1996).

LITERATURE CITED

Arnold, C. D., D. Gerlach, D. Spies, J. A. Matts, Y. A. Sytnikova *et al.*, 2014 Quantitative genome-wide enhancer activity maps for five *Drosophila* species show functional enhancer conservation and turnover during cis-regulatory evolution. Nat. Genet. 46: 685–692. https://doi.org/10.1038/ng.3009

Beye, M., M. Hasselmann, M. K. Fondrk, R. E. Page, and S. W. Omholt, 2003 The gene *csd* is the primary signal for sexual development in the honeybee and encodes an SR-type protein. Cell 114: 419–429. https://doi.org/10.1016/S0092-8674(03)00606-8

Beye, M., C. Seelmann, T. Gempe, M. Hasselmann, X. Vekemans *et al.*, 2013 Gradual molecular evolution of a sex determination switch through incomplete penetrance of femaleness. Curr. Biol. 23: 2559–2564. https://doi.org/10.1016/j.cub.2013.10.070

Björklund, A. K., D. Ekman, S. Light, J. Frey-Skott, and A. Elofsson, 2005 Domain rearrangements in protein evolution. J. Mol. Biol. 353: 911–923. https://doi.org/10.1016/j.jmb.2005.08.067

Bradley, R. K., X. Y. Li, C. Trapnell, S. Davidson, L. Pachter *et al.*, 2010 Binding site turnover produces pervasive quantitative changes in transcription factor binding between closely related *Drosophila* species. PLoS Biol. 8: e1000343. https://doi.org/10.1371/journal.pbio.1000343

Buck, M., A. Burgess, R. Stirzaker, K. Krauer, and T. Sculley, 2006 Epstein-Barr virus nuclear antigen 3A contains six nuclear-localization signals. J. Gen. Virol. 87: 2879–2884. https://doi.org/10.1099/vir.0.81927-0

Cardinal, S., and B. N. Danforth, 2013 Bees diversified in the age of eudicots. Proc. Biol. Sci. 280: 20122686. https://doi.org/10.1098/rspb.2012.2686

Cordeddu, V., E. Di Schiavi, L. A. Pennacchio, A. Ma'ayan, A. Sarkozy *et al.*, 2009 Mutation of SHOC2 promotes aberrant protein N-myristoylation and causes Noonan-like syndrome with loose anagen hair. Nat. Genet. 41: 1022–1026. https://doi.org/10.1038/ng.425

Cormack, B. P., R. H. Valdivia, and S. Falkow, 1996 FACS-optimized mutants of the green fluorescent protein (GFP). Gene 173: 33–38. https://doi.org/10.1016/0378-1119(95)00685-0

Davey, N. E., M. S. Cyert, and A. M. Moses, 2015 Short linear motifs - *ex nihilo* evolution of protein regulation. Cell Commun. Signal. 13: 43. https://doi.org/10.1186/s12964-015-0120-z

Davey, N. E., K. Van Roey, R. J. Weatheritt, G. Toedt, B. Uyar *et al.*, 2012 Attributes of short linear motifs. Mol. Biosyst. 8: 268–281. https://doi.org/10.1039/C1MB05231D

Dermitzakis, E. T., and A. G. Clark, 2002 Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover. Mol. Biol. Evol. 19: 1114–1121. https://doi.org/10.1093/oxfordjournals.molbev.a004169

Di Fiore, B., N. E. Davey, A. Hagting, D. Izawa, J. Mansfeld *et al.*, 2015 The ABBA motif binds APC/C activators and is shared by APC/C substrates and regulators. Dev. Cell 32: 358–372. https://doi.org/10.1016/j.devcel.2015.01.003

Diella, F., N. Haslam, C. Chica, A. Budd, S. Michael *et al.*, 2008 Understanding eukaryotic linear motifs and their role in cell signaling and regulation. Frontiers in Bioscience-Landmark 13: 6580–6603. https://doi.org/10.2741/3175

Dingwall, C., and R. A. Laskey, 1991 Nuclear targeting sequences–a consensus? Trends Biochem. Sci. 16: 478–481. https://doi.org/10.1016/0968-0004(91)90184-W

Dinkel, H., K. Van Roey, S. Michael, N. E. Davey, R. J. Weatheritt *et al.*, 2014 The eukaryotic linear motif resource ELM: 10 years and counting. Nucleic Acids Res. 42: D259–D266. https://doi.org/10.1093/nar/gkt1047

Dyson, H. J., and P. E. Wright, 2005 Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. 6: 197–208. https://doi.org/10.1038/nrm1589

Fagerlund, R., K. Melen, L. Kinnunen, and I. Julkunen, 2002 Arginine/lysine-rich nuclear localization signals mediate interactions between dimeric STATs and importin alpha 5. J. Biol. Chem. 277: 30072–30078. https://doi.org/10.1074/jbc.M202943200

Finn, R. D., P. Coggill, R. Y. Eberhardt, S. R. Eddy, J. Mistry *et al.*, 2016 The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44: D279–D285. https://doi.org/10.1093/nar/gkv1344

Fried, H., and U. Kutay, 2003 Nucleocytoplasmic transport: taking an inventory. Cell. Mol. Life Sci. 60: 1659–1688. https://doi.org/10.1007/s00018-003-3070-3

Fuxreiter, M., P. Tompa, and I. Simon, 2007 Local structural disorder imparts plasticity on linear motifs. Bioinformatics 23: 950–956. https://doi.org/10.1093/bioinformatics/btm035

Gempe, T., M. Hasselmann, M. Schiott, G. Hause, M. Otte *et al.*, 2009 Sex determination in honeybees: two separate mechanisms induce and maintain the female pathway. PLoS Biol. 7: e1000222. https://doi.org/10.1371/journal.pbio.1000222

Goldman, A., J. Roy, B. Bodenmiller, S. Wanka, C. R. Landry *et al.*, 2014 The calcineurin signaling network evolves via conserved kinase-phosphatase modules that transcend substrate identity. Mol. Cell 55: 422–435. https://doi.org/10.1016/j.molcel.2014.05.012

Hasselmann, M., T. Gempe, M. Schiott, C. G. Nunes-Silva, M. Otte *et al.*, 2008a   Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. Nature 454: 519–522. https://doi.org/10.1038/nature07052

Hasselmann, M., S. Lechner, C. Schulte, and M. Beye, 2010   Origin of a function by tandem gene duplication limits the evolutionary capability of its sister copy. Proc. Natl. Acad. Sci. USA 107: 13378–13383. https://doi.org/10.1073/pnas.1005617107

Hasselmann, M., X. Vekemans, J. Pflugfelder, N. Koeniger, G. Koeniger *et al.*, 2008b   Evidence for convergent nucleotide evolution and high allelic turnover rates at the complementary sex determiner gene of Western and Asian honeybees. Mol. Biol. Evol. 25: 696–708. https://doi.org/10.1093/molbev/msn011

Hediger, M., C. Henggeler, N. Meier, R. Perez, G. Saccone *et al.*, 2010   Molecular characterization of the key switch F provides a basis for understanding the rapid divergence of the sex-determining pathway in the housefly. Genetics 184: 155–170. https://doi.org/10.1534/genetics.109.109249

Holt, L. J., B. B. Tuch, J. Villen, A. D. Johnson, S. P. Gygi *et al.*, 2009   Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. Science 325: 1682–1686. https://doi.org/10.1126/science.1172867

Hoshijima, K., K. Inoue, I. Higuchi, H. Sakamoto, and Y. Shimura, 1991   Control of *doublesex* alternative splicing by *transformer* and *transformer-2* in *Drosophila*. Science 252: 833–836. https://doi.org/10.1126/science.1902987

Koch, V., I. Nissen, B. D. Schmitt, and M. Beye, 2014   Independent evolutionary origin of *fem* paralogous genes and complementary sex determination in Hymenopteran insects. PLoS One 9: e91883. https://doi.org/10.1371/journal.pone.0091883

Kosugi, S., M. Hasebe, M. Tomita, and H. Yanagawa, 2009   Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proc. Natl. Acad. Sci. USA 106: 10171–10176. https://doi.org/10.1073/pnas.0900604106

Lechner, S., L. Ferretti, C. Schoning, W. Kinuthia, D. Willemsen *et al.*, 2014   Nucleotide variability at its limit? Insights into the number and evolutionary dynamics of the sex-determining specificities of the honey bee *Apis mellifera*. Mol. Biol. Evol. 31: 272–287. https://doi.org/10.1093/molbev/mst207

Li, H., and P. M. Bingham, 1991   Arginine/serine-rich domains of the su(wa) and tra RNA processing regulators target proteins to a subnuclear compartment implicated in splicing. Cell 67: 335–342. https://doi.org/10.1016/0092-8674(91)90185-2

Ludwig, M. Z., C. Bergman, N. H. Patel, and M. Kreitman, 2000   Evidence for stabilizing selection in a eukaryotic enhancer element. Nature 403: 564–567. https://doi.org/10.1038/35000615

Ludwig, M. Z., N. H. Patel, and M. Kreitman, 1998   Functional analysis of *eve* stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. Development 125: 949–958.

Lynch, M., 2007   The evolution of genetic networks by non-adaptive processes. Nat. Rev. Genet. 8: 803–813. https://doi.org/10.1038/nrg2192

Makkerh, J. P., C. Dingwall, and R. A. Laskey, 1996   Comparative mutagenesis of nuclear localization signals reveals the importance of neutral and acidic amino acids. Curr. Biol. 6: 1025–1027. https://doi.org/10.1016/S0960-9822(02)00648-6

Marfori, M., A. Mynott, J. J. Ellis, A. M. Mehdi, N. F. Saunders *et al.*, 2011   Molecular basis for specificity of nuclear import and prediction of nuclear localization. Biochim. Biophys. Acta 1813: 1562–1577. https://doi.org/10.1016/j.bbamcr.2010.10.013

Mészáros, B., G. Erdos, and Z. Dosztanyi, 2018   IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. 46: W329–W337. https://doi.org/10.1093/nar/gky384

Moses, A. M., and C. R. Landry, 2010   Moving from transcriptional to phospho-evolution: generalizing regulatory evolution? Trends Genet. 26: 462–467. https://doi.org/10.1016/j.tig.2010.08.002

Moses, A. M., M. E. Liku, J. J. Li, and R. Durbin, 2007   Regulatory evolution in proteins by turnover and lineage-specific changes of cyclin-dependent kinase consensus sites. Proc. Natl. Acad. Sci. USA 104: 17713–17718. https://doi.org/10.1073/pnas.0700997104

Moses, A. M., D. A. Pollard, D. A. Nix, V. N. Iyer, X. Y. Li *et al.*, 2006   Large-scale turnover of functional transcription factor binding sites in *Drosophila*. PLOS Comput. Biol. 2: e130. https://doi.org/10.1371/journal.pcbi.0020130

Neduva, V., and R. B. Russell, 2005   Linear motifs: evolutionary interaction switches. FEBS Lett. 579: 3342–3345. https://doi.org/10.1016/j.febslet.2005.04.005

Nguyen Ba, A. N., A. Pogoutse, N. Provart, and A. M. Moses, 2009   NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. BMC Bioinformatics 10: 202. https://doi.org/10.1186/1471-2105-10-202

Nguyen Ba, A. N., B. Strome, J. J. Hua, J. Desmond, I. Gagnon-Arsenault *et al.*, 2014   Detecting functional divergence after gene duplication through evolutionary changes in posttranslational regulatory sequences. PLOS Comput. Biol. 10: e1003977. https://doi.org/10.1371/journal.pcbi.1003977

Pane, A., M. Salvemini, P. D. Bovi, C. Polito, and G. Saccone, 2002   The *transformer* gene in *Ceratitis capitata* provides a genetic basis for selecting and remembering the sexual fate. Development 129: 3715–3725.

Parker, G. E., R. M. Sandoval, H. A. Feister, J. P. Bidwell, and S. J. Rhodes, 2000   The homeodomain coordinates nuclear entry of the Lhx3 neuroendocrine transcription factor and association with the nuclear matrix. J. Biol. Chem. 275: 23891–23898. https://doi.org/10.1074/jbc.M000377200

Ramírez, S. R., J. C. Nieh, T. B. Quental, D. W. Roubik, V. L. Imperatriz-Fonseca *et al.*, 2010   A molecular phylogeny of the stingless bee genus Melipona (Hymenoptera: Apidae). Mol. Phylogenet. Evol. 56: 519–525. https://doi.org/10.1016/j.ympev.2010.04.026

Rizzo, M. A., G. H. Springer, B. Granada, and D. W. Piston, 2004   An improved cyan fluorescent protein variant useful for FRET. Nat. Biotechnol. 22: 445–449. https://doi.org/10.1038/nbt945

Robbins, J., S. M. Dilworth, R. A. Laskey, and C. Dingwall, 1991   Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell 64: 615–623. https://doi.org/10.1016/0092-8674(91)90245-T

Romiguier, J., S. A. Cameron, S. H. Woodard, B. J. Fischman, L. Keller *et al.*, 2016   Phylogenomics controlling for base compositional bias reveals a single origin of eusociality in corbiculate bees. Mol. Biol. Evol. 33: 670–678. https://doi.org/10.1093/molbev/msv258

Schulte, C., G. Leboulle, M. Otte, B. Grunewald, N. Gehne *et al.*, 2013   Honey bee promoter sequences for targeted gene expression. Insect Mol. Biol. 22: 399–410. https://doi.org/10.1111/imb.12031

Südbeck, P., and G. Scherer, 1997   Two independent nuclear localization signals are present in the DNA-binding high-mobility group domains of SRY and SOX9. J. Biol. Chem. 272: 27848–27852. https://doi.org/10.1074/jbc.272.44.27848

Suijkerbuijk, S. J., T. J. van Dam, G. E. Karagoz, E. von Castelmur, N. C. Hubner *et al.*, 2012   The vertebrate mitotic checkpoint protein BUBR1 is an unusual pseudokinase. Dev. Cell 22: 1321–1329. https://doi.org/10.1016/j.devcel.2012.03.009

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei *et al.*, 2011   MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28: 2731–2739. https://doi.org/10.1093/molbev/msr121

Verhulst, E. C., L. W. Beukeboom, and L. van de Zande, 2010   Maternal control of haplodiploid sex determination in the wasp *Nasonia*. Science 328: 620–623. https://doi.org/10.1126/science.1185805

Walther, R. F., E. Atlas, A. Carrigan, Y. Rouleau, A. Edgecombe *et al.*, 2005   A serine/threonine-rich motif is one of three nuclear localization signals that determine unidirectional transport of the mineralocorticoid

receptor to the nucleus. J. Biol. Chem. 280: 17549–17561. https://doi.org/10.1074/jbc.M501548200

Xia, X., 2017 DAMBE6: New Tools for Microbial genomics, phylogenetics, and molecular evolution. J. Hered. 108: 431–437. https://doi.org/10.1093/jhered/esx033

Xia, X., Z. Xie, M. Salemi, L. Chen, and Y. Wang, 2003 An index of substitution saturation and its application. Mol. Phylogenet. Evol. 26: 1–7. https://doi.org/10.1016/S1055-7903(02)00326-3

Yasuhara, N., M. Oka, and Y. Yoneda, 2009 The role of the nuclear transport system in cell differentiation. Semin. Cell Dev. Biol. 20: 590–599. https://doi.org/10.1016/j.semcdb.2009.05.003

Zhang, J., and M. Nei, 1997 Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. J. Mol. Evol. 44: S139–S146. https://doi.org/10.1007/PL00000067

Communicating editor: D. Gresham