J Mol Evol (2002) 54:42–53
DOI: 10.1007/s00239-001-0016-y

JOURNAL OF MOLECULAR EVOLUTION

Maximum Likelihood Methods Reveal Conservation of Function Among Closely Related Kinesin Families

Carolyn J. Lawrence,[1] Russell L. Malmberg,[1] Michael G. Muszynski,[2] R. Kelly Dawe[1,3]

[1] University of Georgia, Department of Botany, Athens, GA 30602, USA
[2] Pioneer Hi-Bred International, Johnston, IA 50131, USA
[3] University of Georgia, Department of Genetics, Athens, GA 30602, USA

Received: 20 February 2001 / Accepted: 5 June 2001

Abstract. We have reconstructed the evolution of the anciently derived kinesin[1] superfamily using various alignment and tree-building methods. In addition to classifying previously described kinesins from protists, fungi, and animals, we analyzed a variety of kinesin sequences from the plant kingdom including 12 from *Zea mays* and 29 from *Arabidopsis thaliana.* Also included in our data set were four sequences from the anciently diverged amitochondriate protist *Giardia lamblia.* The overall topology of the best tree we found is more likely than previously reported topologies and allows us to make the following new observations: (1) kinesins involved in chromosome movement including MCAK, chromokinesin, and CENP-E may be descended from a single ancestor; (2) kinesins that form complex oligomers are limited to a monophyletic group of families; (3) kinesins that crosslink antiparallel microtubules at the spindle midzone including BIMC, MKLP, and CENP-E are closely related; (4) *Drosophila* NOD and human KID group with other characterized chromokinesins; and (5) *Saccharomyces* SMY1 groups with kinesin-I sequences, forming a family of kinesins capable of class V myosin interactions. In addition, we found that one monophyletic clade composed exclusively of sequences with a C-terminal motor domain contains all known minus end-directed kinesins.

Key words: Kinesin — Motor — KLP — KHC — Phylogeny — Phylogenomics

Introduction

The kinesins[1] constitute a functionally diverse superfamily of ATP-dependent microtubule-based motor proteins. Among their many activities are movement of chromosomes (Yen et al. 1992), movement of membrane-bounded organelles (Hall and Hedgecock 1991), regulation of microtubule and spindle pole dynamics (Nislow et al. 1992; Sawin and Mitchison 1991), and assembly and maintenance of flagella (Morris and Scholey 1997). Much of this versatility in function can be attributed to the unique structural properties of several forms of the mature kinesin holoenzyme (Fig. 1A). Some kinesins function as a single heavy chain (Okada and Hirokawa. 1999), but others assemble into homodimeric and heterodimeric forms (Bloom et al. 1988; Cole et al. 1993). In addition, some dimeric forms also bind subunits called kinesin-associated proteins (KAPs) (Manning and Snyder 2000).

Within most kinesin heavy chains are three major domains arranged in a modular fashion; the motor domain, neck, and stalk. Based upon motor position, three structural variants of the heavy chain are known to exist: N-type, I-type, and C-type (Fig. 1B). The motor of N-type variants is located at the amino terminus of the molecule and is followed by the neck and stalk. C-type

Correspondence to: R. Kelly Dawe; *email:* kelly@dogwood.botany.uga.edu

[1] "Kinesin" is used to refer to all kinesin-like proteins and is not limited to the original kinesin discovered in *Loligo pealii.*

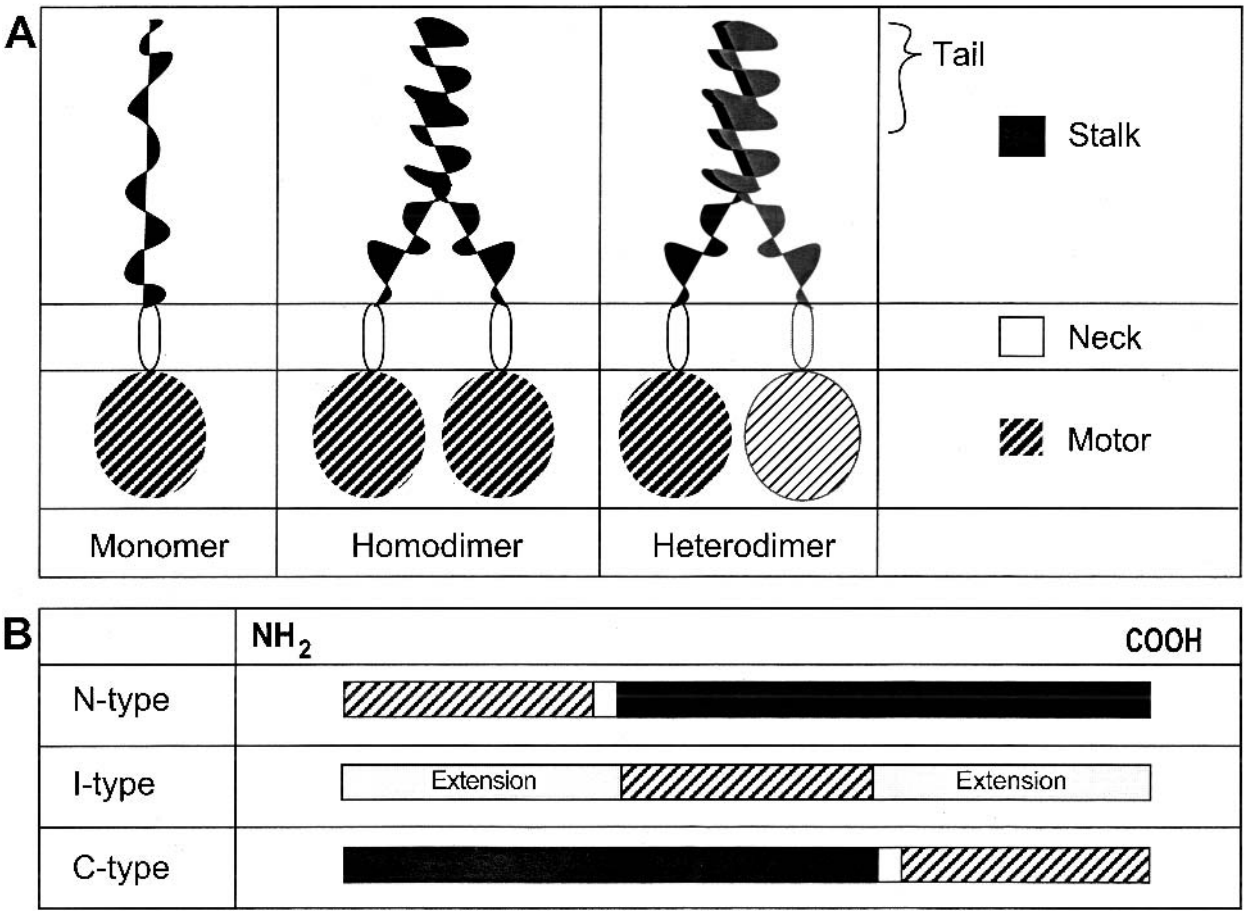


Fig. 1. Kinesin structure. **A** Three types of kinesin exist: monomeric, homodimeric, and heterodimeric. **B** The kinesin heavy-chain arrangement can be N-type, I-type, or C-type.

variants have exactly the opposite order of domains as the N-type variants, putting the motor at the carboxy terminus of the molecule. In I-type variants, no neck domain is present and amino- and carboxy-terminal sequence extensions place the motor in an internal position.

Each domain of the kinesin heavy chain has a defined function. The motor binds microtubules in an ATP-dependent manner (Yang et al. 1989). It is approximately 350 amino acids in length and its sequence is conserved among all variant types. Highly conserved motifs are present at the ATP binding site (FAYGQTGSGKT) and at microtubule binding sites (SSRSH, VDLAGSE, and HIPYR). When attached to a microtubule, the direction that a kinesin moves is dependent upon its neck sequence (Endow and Waligora 1998). Characterized N-type kinesins are plus end-directed and are known to have a neck consisting of approximately 18 amino acids, while characterized C-type kinesins are minus end-directed and have a nonhomologous neck of approximately 14 amino acids (reviewed by Endow 1999). The directional movement of kinesins combined with the fact that microtubules are laid out with respect to microtubule organizing centers (MTOCs) allows for transport to specific compartments of the cell (Fig. 2). Kinesin stalks are comprised of α-helical coiled-coils and are highly variable in both sequence and length. In some kinesins, a terminal portion of the stalk is referred to as the tail and is able to associate with KAPs. This association may be involved in regulation of motor activity and cargo association (Rice et al. 1999; Seiler et al. 2000).

In an effort to relate the structural and functional diversity of the kinesins to their evolutionary history, we carried out a detailed phylogenetic analysis of the kinesin superfamily. Our approach followed that of Goodson et al. (1994), who generated phylogenies based on the conserved motor domain of the kinesin heavy chain. Our data set was made up of 137 kinesin sequences including 69 from animals, 44 from plants, 15 from fungi, and 9 from protists. Many of the proteins have been the subjects of extensive functional analyses. By mapping conserved functions onto monophyletic clades of the kinesin tree, we endeavored not only to provide a new framework for interpreting kinesin origins and evolution, but also to allow informed inferences to be made about the function of kinesins for which only sequence information is known.

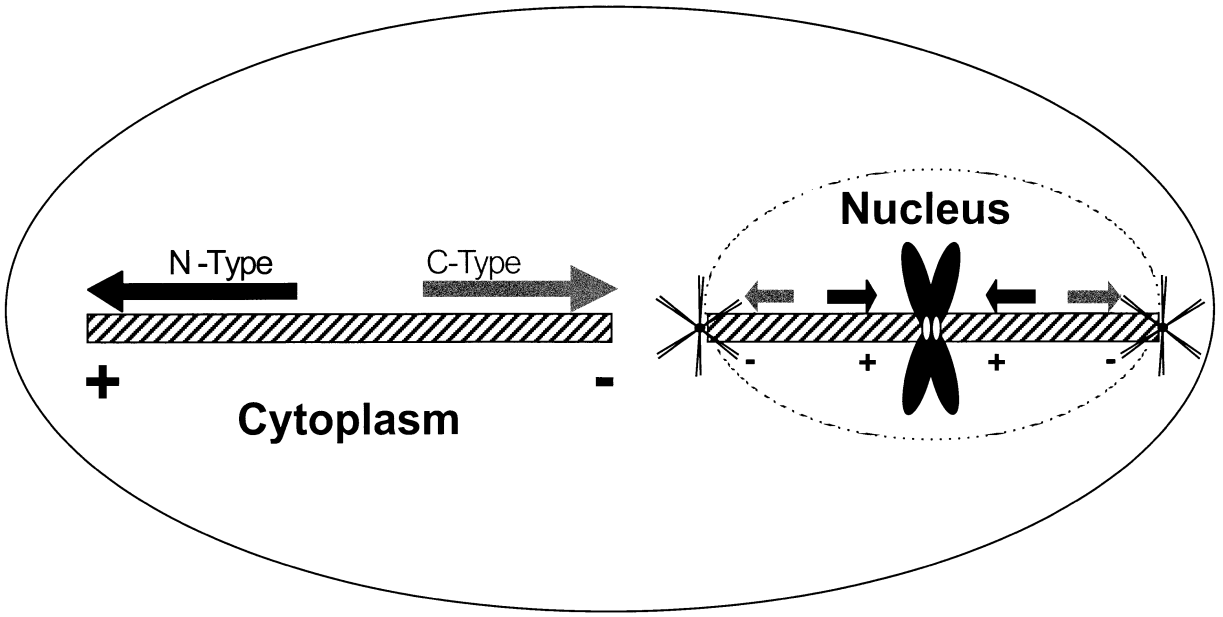


Fig. 2. Kinesin motor directionality. The minus ends of microtubules are relatively stable and tend to be anchored at MTOCs such as centrosomes and basal bodies (animals, protists), spindle pole bodies (fungi), and the nuclear envelope (plants). Plus ends face away from MTOCs and are subject to rapid growth and shrinkage events (reviewed by Inoue and Salmon 1995). In some fungi and protists, the spindle forms within the nuclear envelope (dashed line), while in plants and animals the spindle does not form until after nuclear envelope breakdown. Asters overlapping the nuclear membrane represent spindle MTOCs. During cell division, chromosomes are attached to the plus ends of microtubules that originate from the spindle poles. Characterized kinesins with N-type necks move toward microtubule plus ends, while those with C-type necks move toward microtubule minus ends.

Materials and Methods

Sequence Acquisition

We fully sequenced 12 *Zea mays* kinesin cDNA clones made available by Pioneer Hi-Bred International. The cDNAs of ZmaKIN1, ZmaKIN2, ZmaKIN3, ZmaKIN5, ZmaKIN9, ZmaKIN11, ZmaKIN13, and ZmaKIN16 are partial and lack their 5′ ends. The protein sequences of all the other kinesins analyzed were downloaded from GenBank (Table 1). For all kinesin motor domains included in GenBank release 115, intron positions were determined using an Exon–Intron Database (Saxonov et al. 2000).

Alignments

We constructed three alignments of the kinesin motor domain. The first (EMBL DS43278) is a progressive alignment produced using the Pileup option within GCG's Seqlab and a BLOSUM 30 transition matrix (Feng and Doolittle 1987; GCG 1982–2000; Henikoff and Henikoff 1992). Hidden Markov models (HMMs) were used to generate two other alignments. The second alignment (EMBL DS43279) was generated with HMMER (Eddy 1998) using the adjusted progressive alignment as a profile, and the third alignment (EMBL DS43280) was generated using HMMpro with seven iterations for profile generation (Net-ID, Inc.). Neither the HMMER nor the HMMpro alignment was adjusted manually.

A mask sequence was used to define which regions of the three alignments possessed enough phylogenetic signal to be used in tree-building analyses (Fig. 3). Our mask omitted from tree-building analyses regions of the alignment with less than 15% similarity based on a BLOSUM 30 matrix (Henikoff and Henikoff 1992) where a score of 3 or higher represented similarity. However, where regions of less than 15% similarity occurred within functional domains of the protein and less than 50% of the sequences possessed a gap, the mask was arbitrarily removed.

Tree-Building

Three types of methods were used to infer tree topologies from our alignments: parsimony heuristics, neighbor-joining (NJ) tree-building, and maximum likelihood (ML) heuristics [using both quartet puzzling (Adachi and Hasegawa 1996) and star decomposition (Strimmer and von Haeseler 1996)]. The NJ tree topologies were derived using uncorrected pairwise distances (Swofford 1999). The Jones et al. (1992) model of substitution rates was used to construct ML quartet puzzling and star decomposition topologies, and to calculate ML branch lengths for all tree topologies generated.

Parsimony and NJ analyses were carried out using PAUP* (Swofford 1999), while ML quartet puzzling analyses were carried out using TREE-PUZZLE (Strimmer and von Haeseler 1999). ML star decomposition and nearest-neighbor interchange (NNI) analyses relied upon ProtML from the MOLPHY 2.3 beta 3 package (Adachi and Hasegawa 1996). All tree topologies generated were unrooted, and ML branch lengths were calculated for each topology using ProtML (Adachi and Hasegawa 1996). Likelihoods were calculated and compared using TREE-PUZZLE (Strimmer and von Haeseler 1999). Bootstrap resampling estimates were conducted by the local RELL bootstrap method (LBP) method described by Adachi and Hasegawa (1996).

Comparing Trees

The kinesin sequence sets analyzed by Hirokawa (1998) and Kim and Endow (2000) were downloaded from GenBank and aligned progres-

Table 1. GenBank accession numbers of kinesin heavy chains

Aspergillus nidulans (*Ani*)	
BIMC	M32075
KLPA	X64603
Arabidopsis thaliana (*Ath*)[a]	
KatA	D11371
KatB	D21137
KatC	D21138
KatD	AF080249
KCBP	L40358
KRP125	AC005896
Bombyx mori (*Bmo*)	
KRP	D21206
Caenorhabditis elegans (*Cel*)[b]	
KHC	L19120
KLP3	Z36753
OSM3	D38632
UNC104	M58582
Cylindrotheca fusiformis (*Cfu*)	
DSK1	U51680
Cricetulus griseus (*Cgr*)	
CHO1	X83575
MCAK	U11790
Chlamydomonas reinhardtii (*Cre*)	
FLA10	L33697
KLP1	X78589
Dictyostelium discoideum (Ddi)	
K7	U41289
Drosophila melanogaster (*Dme*)	
KHC	M24441
KLP38B	X99617
KLP3A	AF132186
KLP61F	U01842
KLP67A	U89264
KLP68D	U15974
KLP73	U81788
NCD	X52814
NOD	M36195
Gallus gallus (*Gga*)	
CHRKIN	U18309
Giardia lamblia (*Gla*)[c]	
KIN1	AC046489
	AC040536
	AC030861
KIN2	AC054922
	AC055186
	AC031619
	AC045902
KIN3	AC033214
	AC038066
	AC049203
	AC050899
KIN4	AC048228
	AC035796
	AC048145
Homo sapiens (*Hsa*)	
ATSV	X90840
CENP-E	Z15005
CMKRP	D26361
KHC	X65873
KID	AB017430
KIN2	Y08319
KSP	U37426
MCAK	U63743
MKLP1	X67155
NKHC	U06698

Table 1. Continued

Leishmania chagasi (*Lch*)	
KIN	L07879
Loligo pealii (*Lpe*)	
KHC	J05258
Mus musculus (*Mmu*)	
KHC	X61435
KHCS	L27153
KHCX	U86090
KIF1A	D29951
KIF1B	D17577
KIF2	D12644
KIF3A	D12645
KIF3B	D26077
KIF4	D12646
KIFC1	D49544
KIFC2	D49545
KLP174	Y09632
Morone saxatalis (*Msa*)	
FKIF2	U64819
Neurospora crassa (*Ncr*)	
KHC	L47106
Nectria haematococca (*Nha*)	
KIN1	U86521
Nicotiana tabaccum (*Nta*)	
KRP125	D83711
TCK1	U52078
Rattus norvegicus (*Rno*)	
KRP2	U44979
Saccharomyces cerevisiae (*Sce*)	
CIN8	M90522
KAR3	M31719
KIP1	Z11962
KIP2	Z11963
KIP3	Z72739
SMY1	M69021
Schizosaccharomyces pombe (*Spo*)[d]	
CUT7	X57513
KLP1	U63916
KLP2	AL136235
Strongylocentrotus purpuratus (*Spu*)	
KHC	X56844
KRP85	L16993
KRP95	U00996
Solanum tuberosum (*Stu*)	
KCBP	L46702
Ustilago maydis (*Uma*)	
KIN1	U92844
KIN2	U92845
Xenopus laevis (*Xla*)	
CTK2	U82809
EG5	X54002
EG52	X71864
KCM1	U36485
KCM2	U36486
KLP1	X82012
KLP2	X94082
KLP3	AJ009839
Zea mays (*Zma*)	
KIN1	AF272749
KIN2	AF272750
KIN3	AF272751
KIN4	AF272752
KIN5	AF272753
KIN6	AF272754
KIN8	AF272755

Table 1. Continued

Zea mays (*Zma*)	
KIN9	AF272756
KIN11	AF272757
KIN13	AF272758
KIN15	AF272759
KIN16	AF272760

[a] Plus 24 additional sequences indicated in Fig. 4 by GenBank accession number.
[b] Plus 14 additional sequences indicated in Fig. 4 by GenBank accession number.
[c] Contigs of sequences by GenBank accession number.
[d] Plus 2 indicated in Fig. 4 by GenBank accession number.

sively using the methods they describe. The data set we used to represent the Kim and Endow tree lacked one sequence (*Lycopersicon esculentum* TKRP) since it could not be located in either GenBank or the kinesin home page (http://blocks.fhcrc.org/~kinesin/) web sites. We generated two unrooted tree topologies per alignment: the original published topology and a modified topology that closely resembled the best tree we found with our data set. The modified topology was created by placing orphans as described under Results, performing NNI to situate those orphans within their respective families, and then ordering families by hand to match closely the overall topology of the most likely tree we generated. The likelihoods of the original and modified topologies were calculated for each alignment using TREE-PUZZLE (Strimmer et al. 2000). Finally, we used the Kishino–Hasegawa test (as implemented by TREE-PUZZLE) to determine whether the two topologies were different at the $p < 0.05$ level of significance (Hasegawa and Kishino 1989; Kishino and Hasegawa 1989; Strimmer et al. 2000).

Results

Sampling Tree Space to Find Likely Topologies

We began our study with 137 kinesin homologues, including 12 from maize that have been submitted to GenBank as a part of this study. To generate a distribution of independently derived tree topologies for likelihood comparisons, we analyzed three alignments (progressive, HMMER, and HMMpro) using parsimony heuristics, NJ, ML star decomposition, and ML quartet puzzling. We found that for each of the three alignments, the most likely tree was generated by ML star decomposition followed by NNI.

To search tree space further for likely topologies, we used the three alignments as the basis for another round of NNI with each of the three tree topologies (i.e., nine alignment–topology combinations were analyzed). The most likely tree from this analysis was the result of NNI with the HMMpro alignment/ML star decomposition topology with the progressive alignment. This tree was more likely than all other trees regardless of which alignment was used as the basis for calculating its likelihood score, suggesting that the second round of NNI identified a previously unexplored region of tree space.

User Changes Further Improved the Best ML Star Decomposition Tree

In addition to generating tree topologies by the methods outlined above, we tested our own hypotheses about kinesin evolution by placing ungrouped proteins within a family of similar function or with similar intron positions and phases (Table 2 and Fig. 3). These "user" changes were imposed on the most likely tree from the previous section. User changes that resulted in a better likelihood value were accepted and those resulting in inferior likelihood values were dismissed. When user changes resulted in a tree a with higher likelihood value, the sequence in question also was placed in all other families as a control. No control placements increased the likelihood of the tree.

The three user changes that increased the likelihood of the tree were the placement of SceSMY1 within the kinesin-I family and the placement of DmeNOD and HsaKID among other chromokinesins. After each user placement, a round of NNI was used to situate each of the transplanted sequences optimally within their new families. While SceSMY1 and DmeNOD remained within their respective families, HsaKID swapped to end up with AthAL049655. Since AthAL049655 was also ungrouped and shared two introns with characterized chromokinesins, we placed HsaKID and AthAL049655 as a unit within the chromokinesin family. This placement was an accepted user change, and the following round of NNI moved these two kinesins only within the chromokinesin family, not out of it. This was the most likely tree produced by our analysis and is shown in Fig. 4. [For a general review of kinesin families, the reader is referred to the kinesin home page http://blocks.fhcrc.org/~kinesin/ and Hirokawa (1998).]

To determine how our results compared to models put forth previously, we compared the overall topology of our most likely tree to overall topologies of two other kinesin phylogenies in the literature (Hirokawa 1998; Kim and Endow 2000). As described under Materials and Methods, we first reconstructed the alignments used for the previously published trees (using the cited methods). Next, we created a topology that closely resembled that of our most likely tree (see Fig. 4) by hand. Finally, we compared the likelihoods of the topologies we created to the topologies of the previously published trees using the Kishino–Hasegawa test (Hasegawa and Kishino 1989; Kishino and Hasegawa 1989). This is an appropriate use of the Kishino–Hasegawa test as defined by Goldman et al. (2000). We found that the overall architecture of the tree in Fig. 4 is more likely than the architecture of trees presented by both Hirokawa (1998) and Kim and Endow (2000) at the $p < 0.05$ level of significance (Hasegawa and Kishino 1989; Kishino and Hasegawa 1989).

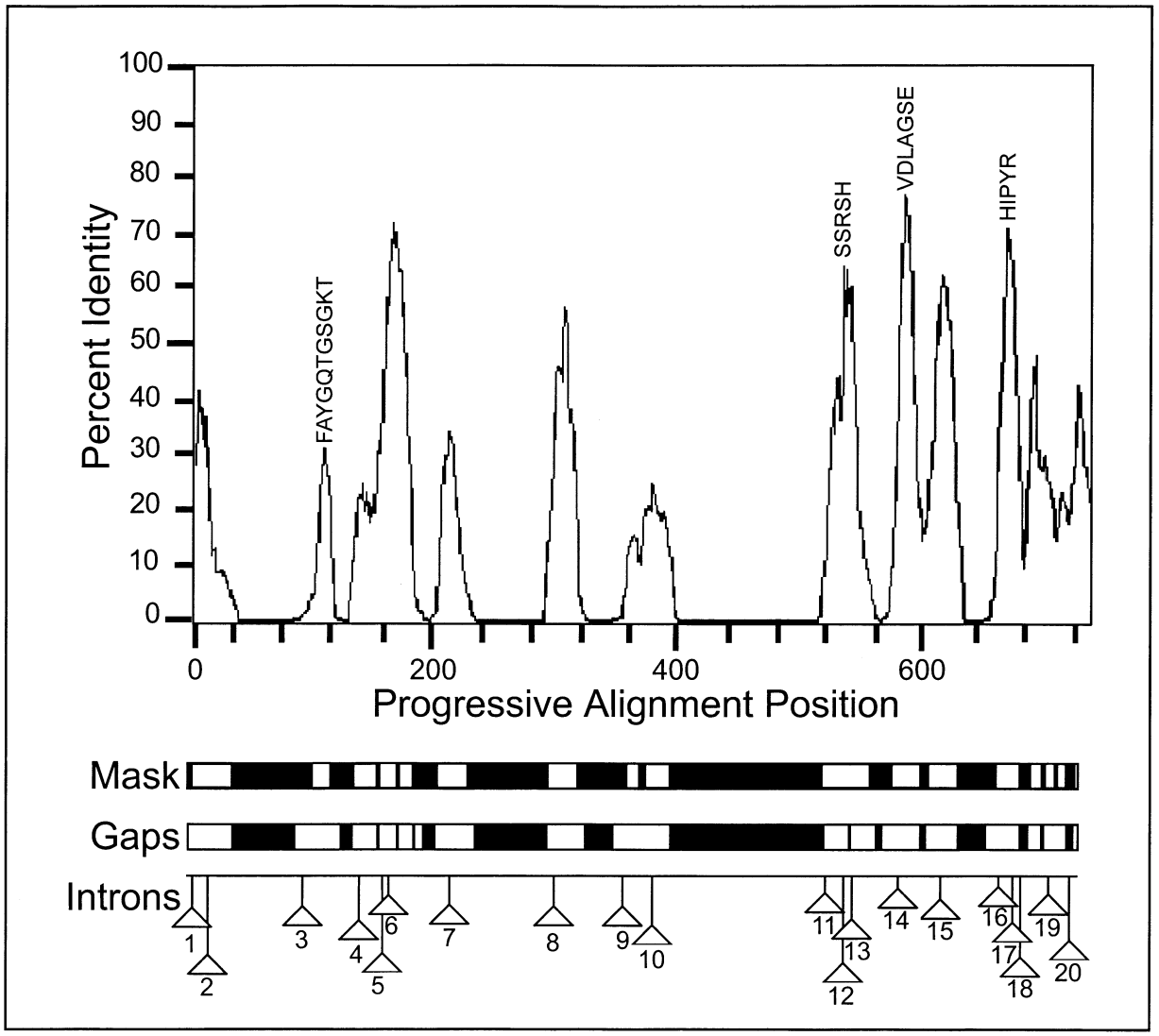


Fig. 3. Percentage identity within a ten-amino acid window along the progressive alignment. *Black areas* along the Mask bar were not included in tree-building analyses. *Black areas* along the Gaps bar indicate regions of the alignment where gaps are present in at least 132 sequences. *Triangles* hanging from the Introns bar represent intron positions relative to the progressive alignment (see also Table 2).

Intron Position Is Conserved in the CENP-E and C-I and C-II Families

In the CENP-E family, members from *Arabidopsis* show intron position and phase conservation (Table 2). Similarly, sequences from *C. elegans* and *Arabidopsis* share intron positions and phases in both the C-I (positions 4 and 9) and the C-II (positions 6, 9, and 18) families. This suggests that introns may have occupied these positions in ancestral C-I and C-II sequences predating the separation of the plant and animal lineages.

Discussion

We begin our discussion with what we believe is the most significant result from our analysis: closely related kinesin families share both structural and functional similarities. We follow with a discussion of how inferences from functional analyses helped us to place orphan kinesins within their respective families and ways in which the methods we used improved upon the details of the tree to resolve a monophyletic origin for kinesins with the C-type domain arrangement. Finally, we appraise the strengths and weaknesses of various alignment and tree-building methods.

Closely Related Kinesin Families Have Similar Functions

The kinesin tree (Fig. 4) can be divided into a large upper clade consisting of the I-type, Kip3, chromokinesin, Unc104, CENP-E, and MKLP families, a large lower clade consisting of the C-type, kinesin-I, and kinesin-II families, and a third clade consisting of only the BIMC

Table 2. Position and phase of introns in the progressive alignment (see Fig. 3)

		Intron position phase																			
Family	Sequence	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Kip3	SpoAL023587	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	SpoZ97211	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	SceKIP3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	CelU53343	0	—	—	0	—	—	—	0	—	—	—	—	—	—	—	1	—	1	—	—
Chromokinesin	CelZ92811	2	—	—	—	0	—	—	—	—	0	—	—	—	—	0	—	—	2	—	—
	AthAL132976	—	—	0	—	0	—	—	0	—	—	—	2	—	—	1	—	0	—	—	—
	CelAL021481	—	—	0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	AthAL049655	—	—	0	0	—	—	—	0	—	—	—	0	—	1	—	—	—	1	—	—
Unc104	CelU23515	—	—	—	0	—	—	1	—	—	—	—	—	—	2	—	—	—	0	—	2
	CelU41536	—	—	2	—	—	—	0	2	—	0	—	—	—	—	2	—	—	—	—	—
	GlaKIN1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
CENP-E	AthAC035679_5	—	—	1	1	—	—	0	—	0	2	—	0	—	—	2	—	—	—	—	—
	AthAC006841_7	—	—	1	1	—	—	0	0	—	—	—	—	1	0	—	—	—	—	—	—
	AthAC002343	—	—	1	1	—	—	0	—	—	0	—	—	0	—	2	—	—	—	—	—
	AthAL109619	—	—	1	1	—	—	0	—	0	—	1	0	—	—	2	—	—	—	—	—
	AthAB028470	—	—	1	1	—	—	0	0	—	1	0	—	0	—	—	2	—	0	—	—
	AthAC006841_14	—	0	—	1	—	—	0	1	0	—	1	—	1	—	0	—	—	0	—	—
	AthAC007727	2	—	1	1	—	0	1	—	—	—	—	0	—	0	0	—	—	0	—	—
	SceKIP2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
MKLP	CelU41007	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	CelU61947	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	AthZ97335	—	—	—	0	0	—	—	2	1	1	0	—	0	—	2	—	—	—	2	—
	GlaKIN3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	CelU61955	—	—	—	—	—	—	2	2	—	—	—	—	—	2	—	—	—	—	0	—
C-I	CelU80450	—	—	—	0	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—
	CelZ81048	—	—	—	0	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—
	CelZ66521	—	—	—	—	—	—	—	—	0	—	—	—	—	—	—	—	—	—	—	—
	DmeNCD	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	AthAF002678	—	—	—	0	—	—	—	0	0	—	—	—	—	0	—	—	—	—	—	—
	AthAF002220	—	—	—	0	—	—	—	0	0	—	—	—	—	0	—	—	—	—	—	—
C-II	AthAC011622	—	—	—	1	—	0	—	—	—	1	—	2	—	—	—	—	—	1	—	—
	AthAC002534	—	—	—	1	—	0	—	—	—	1	—	2	—	—	—	—	—	1	—	—
	AthAC002535	—	—	—	1	—	0	—	0	—	—	—	2	—	—	—	—	—	1	—	—
	AthAC003114	—	—	—	—	—	0	1	—	—	—	—	2	—	—	—	—	—	1	—	—
	AthAC009400_19	—	—	—	1	—	—	0	—	0	—	1	—	0	0	—	2	—	—	—	—
	AthAC009400_8	—	—	—	1	—	0	1	—	—	—	—	2	—	—	—	—	—	1	—	—
	AthAC006340	—	—	—	1	—	—	—	—	2	—	—	2	—	—	—	2	—	—	—	—
	AthAC005223	—	—	—	1	—	0	—	—	2	—	—	2	—	—	—	2	—	1	—	—
	CelKLP3	—	—	—	—	—	0	—	—	2	—	0	—	1	—	—	—	—	1	—	—
Kinesin-I	NhaKIN1	—	—	0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Kinesin-II	CelZ68161	0	2	—	—	—	—	—	0	—	0	2	—	—	—	—	—	—	—	0	—
	CelOSM3	1	—	0	0	—	—	—	0	—	—	0	—	—	—	—	—	0	—	—	—
BIMC	CelZ77659	2	—	—	—	2	—	2	—	—	—	—	—	—	—	—	—	—	—	—	—
	AthKRP125	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—
	AthAC007171	2	—	—	0	—	0	—	—	—	—	—	—	—	—	0	2	—	—	—	—
	SceCIN8	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	SceKIP1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	SpoCUT7	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	GlaKIN2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Ungrouped	AthZ97336	—	—	—	—	—	—	—	—	0	—	—	0	—	—	0	—	0	—	—	—
	CelZ78201	—	0	—	—	—	1	2	—	—	—	—	—	1	—	—	2	—	—	—	—
	GlaKIN4	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

family. The members of families in the upper clade function as simple monomers or homodimers and most share a common function: the movement of chromosomes. For instance, within the I-type and CENP-E families, CgrMCAK and HsaCENP-E localize to different regions of the kinetochore, where they are necessary for proper movement of chromosomes during cell division (Wordeman and Mitchison 1995; Yen et al. 1992). In addition, most characterized members of the chromokinesin family bind DNA directly to participate in chromosome movement (Fig. 4) (Vernos et al. 1995; Wang and Adler 1995). In the Unc104 family, the basal-most characterized family member, DmeKLP38B, functions in chromosome condensation and chromosome movement but is hypoth-

esized to exist as a monomer like other Unc104 family members (Alphey et al. 1997; Molina et al. 1997). This suggests that the monomeric motors of the Unc104 family, most of which are involved in transporting vesicles and mitochondria (Dorner et al. 1998; Hall and Hedgecock 1991), may have evolved from an ancestor that associated with chromosomes.

In contrast to the simple structure and chromosome-oriented functions of the upper clade, members of the lower clade demonstrate complex oligomerization events and have an array of different functions. In these families, the holoenzyme is formed by oligomerization of KAPs with heavy chain dimers (Manning and Snyder 2000). In the C-I family, SceKAR3 interacts with two KAPs that function to target SceKAR3 differentially within the cell: SceKAR3 with the KAP SceCIK1 shows diffuse localization within the nucleus, while SceKAR3 bound to SceVIK1 localizes to the spindle pole body (Manning et al. 1999). The KAPs of kinesin-I family members, called kinesin light chains, are expressed as multiple isoforms which target kinesin to different organelles (Gyoeva et al. 2000). Kinesin-II family members are heterotrimeric consisting of a heavy-chain heterodimer bound to a single KAP, which, like kinesin light chains, may function to mediate cargo specificity (Cole 1999a; Wedaman et al. 1996). Likewise, the lowermost clade within the C-I family is made up of kinesins that bind calmodulin light chains in the presence of Ca^{2+} to inhibit motor activity (Deavours et al. 1998; Narasimhulu and Reddy 1998; Vos et al. 2000). While other proteins have been shown to associate with various kinesins, many are yet uncharacterized, and others appear to interact only transiently or as cargo (reviewed by Manning and Snyder 2000).

In addition to the structural and functional conservation within the large upper and lower clades, a third functional group of closely related families exists: families that mediate antiparallel microtubule interactions. The BIMC, MKLP, and CENP-E families form this group. BIMC family members like AniBIMC are homotetrameric and are able to crossbridge antiparallel microtubules to form and maintain the bipolar mitotic spindle. They localize to the spindle throughout cell division, becoming concentrated at the spindle midzone during anaphase B (Sawin et al. 1992). The animal kinetochore protein CENP-E also relocalizes to the spindle midzone beginning at anaphase onset, as do several animal MKLP family members (Boleti et al. 1996; Yao et al. 1997).

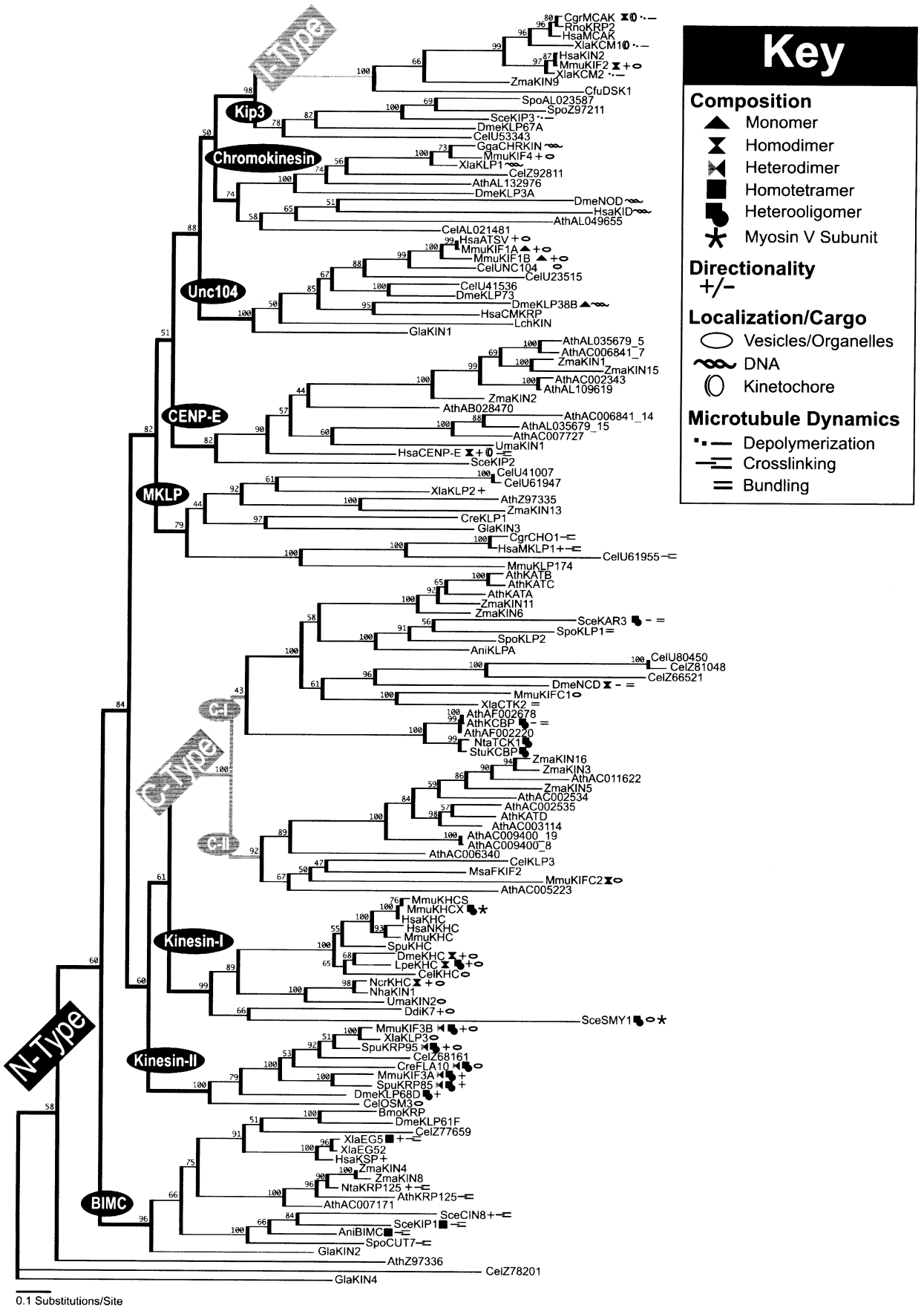
"Orphan" Kinesins Find a Family

Our analysis placed three sequences SceSMY, DmeNOD, and HsaKID within kinesin families with similar functions, a result not found by Hirokawa (1998) or Kim and Endow (2000). In our most likely tree, SceSMY1 appears to be a divergent member of the kinesin-I family. This finding is supported by the fact that both SceSMY1 and MmuKHCX interact with class V myosins (Beningo et al. 2000; Huang et al. 1999). In addition, the tails of SceSMY1 and MmuKHCX share a small region of sequence similarity near the ends of their respective coiled-coil stalks (Beningo et al. 2000). We also found that DmeNOD and HsaKID, both of which are known chromokinesins (Afshar et al. 1995; Tokai et al. 1996), group with other chromokinesins. Interestingly, we find that there are two distinct groups within the chromokinesin family: a lower clade with DmeNOD and HsaKID and an upper clade with GgaCHRKIN and XlaKLP1. This grouping strengthens the hypothesis made by Heald (2000) and Funibiki and Murray (2000) that two distinct chromokinesin classes exist. These authors argue that there is a KLP1/CHRKIN class that organizes spindles around chromosomes and a KID/NOD class that is involved in moving chromosomes toward the metaphase plate (Funibiki and Murray 2000; Heald 2000). As mentioned earlier, DmeKLP38B of the Unc104 family also interacts with chromosomes directly (Alphey et al. 1997; Molina et al. 1997). We hypothesize that DmeKLP38B may represent a third class of chromokinesin that is closely related to the monomeric Unc104 kinesins.

Although HsaCENP-E was referred to as an orphan in previously published kinesin phylogenies (Goodson et al. 1994; Hirokawa 1998; Kim and Endow 2000; Moore and Endow 1996), we find that it is a member of a monophyletic group consisting of 14 sequences. We call this group the CENP-E family after its best-characterized member, HsaCENP-E. HsaCENP-E accumulates in the cytoplasm during the G2 phase of the cell cycle, then relocalizes to the fibrous corona of kinetochores as soon as the nuclear envelope breaks down during cell division (Yen et al. 1992). It remains bound to kinetochores from congression through metaphase, then at anaphase leaves the kinetochores and localizes to the spindle midzone (Yao et al. 1997; Yen et al. 1992). HsaCENP-E is required for chromosomes to align at the metaphase plate and may function to tether kinetochores to dynamic microtubule plus ends (Lombillo et al. 1995; Wood et al. 1997). No plant CENP-E family members have been functionally characterized, and the UmaKIN I deletion mutant has no discernible phenotype (Lehmler et al. 1997). The only other functionally characterized protein in the CENP-E clade is SceKIP2. SceKIP2 participates in nuclear migration and functions to stabilize cytoplasmic microtubules (Miller et al. 1998), suggesting that, like HsaCENP-E, SceKIP2 function may be restricted to microtubule plus ends.

The Evolution of Minus-End Directionality

Domain swapping experiments indicate that the neck domain of kinesin dictates its direction of movement (Case


Key
Composition
Monomer
Homodimer
Heterodimer
Homotetramer
Heterooligomer
Myosin V Subunit
Directionality
+/−
Localization/Cargo
Vesicles/Organelles
DNA
Kinetochore
Microtubule Dynamics
Depolymerization
Crosslinking
Bundling
N-Type
C-Type
BIMC
Kinesin-II
Kinesin-I
MKLP
CENP-E
Unc104
Chromokinesin
Kip3
C-I
C-II
0.1 Substitutions/Site

et al. 1997; Endow and Waligora 1998; Sablin et al. 1998). Our phylogenetic data suggest that the sequence of the kinesin motor domain also predicts (but does not confer) directionality, since all C-type kinesins used in our analysis grouped together (Fig. 4). Our data also indicate that the first kinesins were probably plus end directed. We base this hypothesis on the fact that in all the kinesin trees we generated, the midpoint fell among kinesins with an N-type domain arrangement. No alignment or method of tree construction yielded a tree with C-type families at the tree's midpoint.

There are two families of C-type kinesins. Both possess plant and animal members, suggesting that the C-type kinesins had diverged into two distinct groups prior to the divergence of plants from animals and fungi. The fact that *C. elegans* and *Arabidopsis* share intron position and phase in both the C-I (introns 4 and 9) and the C-II (introns 6, 9, and 18) families supports this assertion (Table 2). Too few C-II family members have been characterized to ascribe a conserved function to them, but the available data suggest that the C-I kinesins function to bundle microtubules during meiosis and mitosis (Fig. 4). Interestingly, the plant KCBPs of the C-I family seem to be an evolutionary intermediate between the N-type and the C-type kinesins. The KCBPs have a C-type neck but also have a carboxy-terminal extension homologous to the N-type neck (data not shown). The fact that AthKCBP is minus end directed (Song et al. 1997) despite the presence of an N-type neck sequence supports Endow's observation that the presence of the C-type neck is sufficient to impose minus-end directedness on the protein (Endow and Waligora 1998).

The Strengths of Using the Likelihood Optimality Criterion

Kuhner and Felsenstein (1994), and later Takahashi and Nei (2000), demonstrated that distance methods such as NJ sometimes outperform ML heuristic search methods. With this in mind, we used various optimality criteria to generate numerous tree topologies and then chose from that set a single most likely tree for each alignment. No parsimony, NJ, or ML quartet puzzling tree produced was more likely than the trees produced by ML star decomposition followed by NNI. However, the tree with the most likely topology overall was produced by taking intron position and experimental observations into account, then placing individual kinesins within the tree by hand (Fig. 4). Not only was this tree more likely than any other tree we produced, but its overall architecture is more likely than the architectures of trees produced by both Hirokawa (1998) and Kim and Endow (2000). Finally, we note that three alignment methods were used in this study. The most likely tree was generated using both the HMMpro and progressive alignments at different stages, suggesting that each alignment had phylogenetic information that was lacking in the others.

There are clear similarities among our best kinesin tree and those most recently published [i.e., Fig. 4 and the Hirokawa (1998) and Kim and Endow (2000) trees]. For instance, the major kinesin families tend to have roughly the same members no matter which methods were used, and all of the trees found a close relationship between the BIMC an the kinesin-II families and among the Unc104, chromokinesin, and I-type families. However, while the Hirokawa tree grouped all C-type kinesins together, the Kim and Endow tree did not. Neither the Hirokawa nor the Kim and Endow trees grouped SceSMY, HsaKID, and DmeNOD with their respective families, and the I-type and Kip3 families, which share the function of microtubule depolymerization, were not grouped together in either of their trees. We believe that the tree we present here not only correlates well with the trees published both by Hirokawa and by Kim and Endow, but also sheds light on some previously unrealized details of kinesin evolution.

Acknowledgments. This research was supported by a grant from the National Science Foundation (MCB9513556) to R.K.D. and an Energy Biosciences grant (DEF602-97ER20286) from the Department of Energy to R.L.M. Additional support was provided to C.J.L. by a National Science Foundation interdisciplinary research training grant (BIR9220329).

Fig. 4. Most likely phylogenetic tree of the kinesin superfamily. The most likely tree that we produced is a product of star decomposition and NNI (Adachi and Hasegawa 1996) with the HMMpro alignment followed by user changes and a second round of NNI with the progressive alignment ($-\ln L = 53{,}007.62$ with the progressive alignment see Results for details). The tree is arbitrarily rooted by GlaKIN4. Numbers at nodes represent the bootstrap probability by the RELL method (Adachi and Hasegawa 1996). Mapped onto the tree are the organizational and functional characteristics of representative sequences (Eisen 1998). The color of branches leading to individual families indicate the heavy-chain domain arrangement (N-type are blue, C-type are red, and I-type are orange). Sequence names are colored to indicate to which kingdom organisms belong (protists are turquoise, plants are green, fungi are purple, and animals are red). Symbols for functional characteristics are plotted to the right of sequence names (set boxed key to the meaning of symbols). We made decisions on how to categorize the functional characteristics of the kinesin superfamily by consulting recently published literature, e.g., monomer (Okada and Hirokawa 1999), homodimer (Maney et al. 1998), heterodimer (Cole 1900b), homotetramer (Kashina et al. 1997), heterooligomer (Manning and Snyder 2000), myosin V (Beningo et al. 2000), directionality (Endow 1999), vesicles/organelles (Hirokawa 1998), DNA (Funibiki and Murray 2000; Heald 2000), kinetochore (Walczak et al. 1996), depolymerization (Desai et al. 1999; Walczak et al. 1996), cross-linking (Kashina et al. 1997; Liao et al. 1994), and bundling (Kao et al. 2000).

References

Adachi J, Hasegawa M (1996) MOLPHY: Programs for molecular phylogenetics based on maximum likelihood, version 2.3. Institute of Statistical Mathematics, Tokyo

Afshar K, Scholey J, Hawley RS (1995) Identification of the chromosome localization domain of the *Drosophila* nod kinesin-like protein. J Cell Biol 131:833–843

Alphey L, Parker L, Hawcroft G, Guo Y, Kaiser K, Morgan G (1997) KLP38B: A mitotic kinesin-related protein that binds PPI. J Cell Biol 138:395–409

Beningo KA, Lillie SH, Brown SS (2000) The yeast kinesin-related protein Smylp exerts its effects on the class V myosin Myo2p via a physical interaction. Mol Biol Cell 11:691–702

Bloom GS, Wagner MC, Pfister KK, Brady ST (1988) Native structure and physical properties of bovine brain kinesin and identification of the ATP-binding subunit polypeptide. Biochemistry 27:3409–3416

Boleti H, Karsenti E, Vernos I (1996) XkIp2, a novel *Xenopus* centrosonial kinesin-like protein required for centrosome separation during mitosis. Cell 84:49–59

Case RB, Pierce DW, Hom-Booher N, Hart CL, Vale RD (1997) The directional preference of kinesin motors is specified by an element outside of the motor catalytic domain. Cell 90:959–966

Cole D (1999a) Kinesin-II, the heterodimeric kinesin. Cell Mol Life Sci 56:217–226

Cole D (1999b) Kinesin-II, coming and going. J Cell Biol 147:463–466

Cole DG, Chinn SW, Wedaman KP, Hall K, Vuong T, Sholey JM (1993) Novel heterotrimeric kinesin-related protein purified from sea urchin eggs. Nature 355:268–270

Deavours BE, Reddy ASN, Walker RA (1998) Ca^{2+}/calmodulin regulation of the *Arabidopsis* kinesin-like calmodulin-binding protein. Cell Motil Cytoskeleton 40:408–416

Desai A, Berma S, Mitchison TJ, Walczak CE (1999) Kin I kinesins are microtubule-destabilizing enzymes. Cell 96:69–78

Dorner C, Ciossek T, Muller S, Moller PH, Ullrich A, Lammers R (1998) Characterization of KIF1C, a new kinesin-like protein involved in vesicle transport from the Golgi apparatus to the endoplasmic reticulum. J Biol Chem 273:20267–20275

Eddy S (1998) HMMER, version 2.1.1. Washington University, St. Louis, MO

Eisen JA (1998) Phylogenomics: Improving functional predictions for uncharacterized genes by evolutionary analysis. Genome Res 8: 163–167

Endow SA (1999) Determinants of molecular motor directionality. Nature Cell Biol 1:163–167

Endow SA, Waligora KW (1998) Determinants of kinesin motor polarity. Science 291:1200–1202

Feng DF, Doolittle RF (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J Mol Evol 25:351–360

Funibiki H, Murray AW (2000) The *Xenopus* chromokinesin Xkid is essential for metaphase chromosome alignment and must be degraded to allow anaphase chromosome movement. Cell 102:411–424

GCG (1982–2000) Wisconsin package, version 10.1. Genetics Computer Group, Inc., Madison

Goldman J, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. Syst Biol 49:651–670

Goodson HV, Kang, SJ, Endow SA (1994) Molecular phylogeny of the kinesin family of microtubule motor proteins. J Cell Sci 107:1875–1884

Gyoeva FK, Bybikova EM, Minin AA (2000) An isoform of kinesin light chain specific for the Golgi complex. J Cell Sci 113:2047–2054

Hall D, Hedgecock E (1991) Kinesin-related gene unc-104 is required for axonal transport of synaptic vesicles in *C. elegans.* Cell 65:837–847

Hasegwwa M, Kishino H (1989) Confidence limits on the maximum-likelihood estimate of the hominoid tree from mitochondrial-DNA sequences. Evolution 43:672–677

Heald R (2000) Motor function in the mitotic spindle. Cell 102:399–402

Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci USA 89:10915–10919

Hirokawa N (1998) Kinesins and dynein superfamily proteins and the mechanism of organelle transport. Science 279:519–526

Huang J, Brady S, Richards B, Stenolen D, Resau J, Copland N, Jenkins N (1999) Direct interaction of microtubule- and actin-based transport motors. Nature 397:267–270

Inoue S, Salmon ED (1995) Force generation by microtubule assembly/disassembly in mitosis and related movements. Mol Biol Cell 6: 1619–1640

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. CABIOS 8:275–282

Kao Y, Deavours BE, Phelps KK, Walker RA, Reddy ASN (2000) Bundling of microtubules by motor and tail domains of a kinesin-like calmodulin-binding protein from *Arabidopsis:* Regulation by Ca^{2+}/calmodulin. Biochem Biophys Res Commun 267:201–207

Kashina AS, Rogers GC, and Scholey JM (1997) The bimC family of kinesins: Essential bipolar mitotic motors driving centrosome separation. Biochim. Biophys Acta 1357:257–271

Kim A, Endow S (2000) A kinesin family tree. J Cell Sci 113:3681–3682

Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J Mol Evol 29:170–179

Kuhner M, Felsenstein J (1994) A simulation comparision of phylogeny algorithms under equal and unequal evolutionary rates. Mol Biol Evol 11:459–468

Lehmler C, Steinberg G, Snetselaar KM, Schliwa M, Kahmann R, Bolker M (1997) Identification of a motor protein required for filamentous growth in *Ustilago maydis.* EMBO J 16:3464–3473

Liao H, Li G, Yen TJ (1994) Mitotic regulation of microtubule cross-linking activity of CENP-E kinetochore protein. Science 265:394–398

Lombillo V, Nislow T, Yen J, Gelfand V, McIntosh J (1995) Antibodies to the kinesin motor domain and CENP-E inhibit depolymerization-dependent motion of chromosomes in vitro. J Cell Biol 128:107–115

Maney T, Hunter A, Wagenbach M, Wordeman (1998) Mitotic centromere-associated kinesin is important for anaphase chromosome segregation. J Cell Biol 142:787–801

Manning BE, Snyder M (2000) Drivers and passengers wanted! The role of kinesin-associated proteins. Trends Cell Biol 10:281–289

Manning BD, Barrett JG, Wallace JA, Granok H, Snyder M (1999) Differential regulation of the Kar3p kinesin-related protein by two associated proteins, Cik1p and Vik1p. J Cell Biol 144:1219–1333

Miller RK, Heller KK, Frisen L, Wallack DL, Loayza D, Gammie AE, Rose MD (1998) The kinesin-related proteins Kip2p and Kip3p, function differently in nuclear migration in yeast. Mol Biol Cell 9:2051–2068

Molina I, Baars S, Brill J, Hales K, Fuller M, Ripoll P (1997) A chromatin-associated kinesin-related protein required for normal mitotic chromosome segregation in *Drosophila.* J Cell Biol 139: 1361–1371

Moore JD, Endow SA (1996) Kinesin proteins: A phylum of motors for microtubule-based motility. BioEssays 18:207–219

Morris L, Scholey J (1997) Heterotrimeric kinesin-II is required for the assembly of motile 9+2 ciliary axonemes on sea urchin embryos. J Cell Biol 138:1009–1022

Narasimhulu SB, Reddy ASN (1998) Characterization of microtubule binding domains in the *Arabidopsis* kinesin-like calmodulin binding protein. Plant Cell 10:957–965

Net-ID, Inc. HMMpro, version 2.2. http://www.netid.com/html/hmmpro.html

Nislow C, Lombillo VA, Kuriyama R, McIntosh JR (1992) A plus-end-directed motor enzyme that moves antiparallel microtubules in vitro localizes to the interzone of mitotic spindles. Nature 359:543–547

Okada Y, Hirokawa N (1999) A processive single-headed motor: Kinesin superfamily protein KIF1A. Science 283:1152–1157

Rice S, Lin AW, Safer D, Hart CL, Naber N, Carragher BO, Cain SM, Pechatnikova E, Wilson-Kubalek EM, Whittaker M, Pate E, Cooke R, Taylor EW, Milligan RA, Vale RD (1999) A structural change in the kinesin motor protein that drives motility. Nature 402:778–784

Sablin EP, Case RB, Dai SC, Hart CL, Ruby A, Vale RD, Fletterick RJ (1998) Direction determination in the minus-end-directed kinesin motor ncd. Nature 395:813–816

Sawin KE, Mitchison TJ (1991) Poleward microtubule flux mitotic spindles assembled in vitro. J Cell Biol 112:941–954

Sawin KE, LeHuellec K, Philippe M, Mitchison TJ (1992) Mitotic spindle organization by a plus-end-directed microtubule motor. Nature 359:540–543

Saxonov S, Daizadeh I, Federov A, Gilbert W (2000) EID: The Exon-Intron Database—An exhaustive database of protein-coding intron-containing genes. Nucleic Acids Res 28:185–190

Seiler S, Kirchner J, Horn C, Kallipolitou A, Woehlke G, Schliwa M (2000) Cargo binding and regulatory sites in the tail of fungal conventional kinesin. Nature Cell Biol 2:333–338

Song H, Golovkin M, Reddy AS, Endow SA (1997) In vitro motility of AtKCBP, a calmodulin-binding kinesin protein of *Arabidopsis*. Proc Natl Acad Sci USA 94:322–327

Strimmer K, von Haeseler A (1996) Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. Mol Biol Evol 13:964–969

Strimmer K, von Haeseler A (1999) PUZZLE, 4.0.2. http://oscar.gen.tcd.ie/ss2000/PUZZLE.html

Strimmer K, Schmidt HA, Vingron M, von Haeseler A (2000) TREE-PUZZLE, version 5.0. http://www.tree-puzzle.de

Swofford D (1999) PAUP*, version 4.Ob2. Sinauer Associates, Sunderland, MA

Takahashi K, Nei M (2000) Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. Mol Biol Evol 17:1251–1258

Tokai N, Fujimoto-Nishiyama A, Toyoshima Y, Tsukita S, Inoue J, Yamamota T (1996) Kid, a novel kinesin-like DNA binding protein, is localized to chromosomes and the mitotic spindle. EMBO J 15:457–467

Vernos I, Raats J, Hirano T, Heasman J, Karsenti E, Wylle C (1995) Xklp1, a chromosomal *Xenopus* kinesin-like protein essential for spindle organization and chromosome positioning. Cell 81:117–127

Vos JW, Safadi F, Reddy AS, Hepler PK (2000) The kinesin-like calmodulin binding protein is differentially involved in cell division. Plant Cell 12:979–990

Walczak C, Mitchison T, Desai A (1996) XKCM1: a *Xenopus* kinesin-related protein that regulates microtubule dynamics during mitotic spindle assembly. Cell 84:37–47

Wang SZ, Adler R (1995) Chromokinesin: A DNA-binding, kinesin-like nuclear protein. J Cell Biol 128:761–768

Wedaman KP, Meyer DW, Rashid DJ, Cole DG, Scholey JM (1996) Sequence and submolecular localization of the 115-kD accessory subunit of the heterotrimeric kinesin-II (KRP85/95) complex. J Cell Biol 132:371–380

Wood KW, Sakowicz R, Goldstein LSB, Cleveland DW (1997) CENP-E is a plus end-directed kinetochore motor required for metaphase chromosome alignment. Cell 91:357–366

Wordeman L, Mitchison T (1995) Identification and partial characterization of mitotic centromere-associated kinesin, a kinesin-related protein that associates with centromeres during mitosis. J Cell Biol 128:95–104

Yang JT, Laymon RA, Goldstein LS (1989) A three-domain structure of kinesin heavy chain revealed by DNA sequence and microtubule binding analyses. Cell 56:879–998

Yao W, Anderson KL, Cleveland DW (1997) The microtubule-dependent motor centromere-associated protein E (CENP-E) is an integral component of kinetochore corona fibers that link centromeres to spindle microtubules. J Cell Biol 139:435–447

Yen T, Gang L, Schaar B, Szilak I, Cleveland D (1992) CENP-E is a putative kinetochore motor that accumulates just before mitosis. Nature 359:536–539