FILE S2

Functional inferences from the analyses of *Dntf-2* and *ran* retrogene protein sequences

Heterospecific complexes formed between rat NTF-2 protein and canine Ran protein have been crystallized, and their mutual interactions determined (BERMAN *et al.* 2002). Additional analysis and crystal structures reveal the interactions of NTF-2 and Ran with other nuclear transport proteins (BULLOCK *et al.* 1996; ISGRO and SCHULTEN 2007; MARCHLER-BAUER *et al.* 2007; MATSUURA and STEWART 2004; RENAULT *et al.* 2001; SEEWALD *et al.* 2002; STEWART *et al.* 1998; VETTER *et al.* 1999). We threaded the protein sequences of the *Dntf-2* and *ran* retrogenes onto the structures of the crystallized paralogs in an effort to infer structural conflicts that might prohibit known parental protein-protein interactions. These analyses were performed using PyMOL software (http://www.pymol.org).

Ntf-2 and *ran* proteins are well known and conserved proteins (QUIMBY *et al.* 2000) that physically interact with each other and play a central role in the transport of proteins to the nucleus (RIBBECK *et al.* 1998). Ran exists in GDP bound inactive form and

GTP bound active form. RanGDP predominantly localizes in the cytoplasm and RanGTP in the nucleus. NTF-2 is a dimer that interacts with nucleoporins and RanGDP during transport to the nucleus (QUIMBY *et al.* 2000). RanGDP also interacts with RanGEF in the nucleus during RanGDP to RanGTP conversion (ISGRO and SCHULTEN 2007), with exportins in the nucleus in order to transport complexes out of the nucleus (KUSANO *et al.* 2003; MATSUURA and STEWART 2004), with Importin β during Importin β's return trip to the cytoplasm (ISGRO and SCHULTEN 2007), and with RanGAP in the cytoplasm during RanGTP hydrolysis to RanGDP (KUSANO *et al.* 2002). Together, these proteins maintain a gradient of RanGDP/RanGTP that is important for protein import and export. Details about the particular residues of NTF-2 and Ran that are known to be involved in protein-protein interactions are given in Figure 2 and File S1.

We have threaded proteins from *Dntf-2*, *Dntf-2* derived retrogenes, *ran*, and *ran* derived retrogenes onto rat NTF-2 and canine Ran, respectively, to identify possible changes in function in the newly duplicated proteins (see Materials and Methods). We know that *Dntf-2* is a gene under purifying selection in Drosophila (K_A/K_S =0.0188; see above), but *ran* is under stronger purifying selection (K_A/K_S =0.0065; see above), likely due to the fact that it carries multiple functions, as discussed above. The selection imposed on Ran relative to DNTF-2 is evident in the alignments of these Drosophila proteins with mammalian orthologs (Figure 2 and File S1).

By threading the *Dntf-2* retrogene proteins (and DNTF-2) from the three Drosophila lineages onto a known NTF-2 crystal structure, we observed that the amino acids involved in interacting with RanGDP and the FxFG repeats of the nucleoporins are conserved or have changes that do not seem to structurally exclude binding to RanGDP or to nucleoporins (File S2, panel A). Even for Da_NTF-2R, the most divergent retrogene protein, the residues important for function are conserved or do not seem to impose overt structural conflicts, suggesting that the retrogenes of Dntf-2 encode transport competent proteins capable of carrying RanGDP across the nuclear membrane.

Similar threading of the *Drosophila ran* retrogene proteins (and Ran) onto the known Ran crystal structure (Figure 2 and File S1, panel B) showed that the proteins encoded by the retrogenes in *D. ananassae* and *D. grimshawi* lineages show little divergence from their parental genes. Most of the amino acids that are important for function are identical or underwent conservative changes that do not appear to impose structural conflicts, suggesting similar functions between Ran and its

retrogenes in these lineages. While *Da_ran-like* and *Dg_ran-like* proteins are similar to the parental proteins, this similarity is not due to a recent origin as the *K*S is saturated between parental and retrogenes. The *K*S is 3.3193 for *Da_ran-like* and is 1.5144 for *Dg_ran-like* calculated using PAML. Accordingly, the *K*A/*K*S ratios are 0.0408 for *Da_ran-like* and 0.0348 for *Dg_ran-like* (Table S4).

The Ran-like in the *D. melanogaster* subgroup is more diverged, and our threading analyses indicate that some changes are likely to abolish some protein-protein interaction interfaces while possibly retaining others (Figure 2). The interface with DNTF-2 appears to be present, as most of the interacting amino acids are identical or show conservative changes. We also posit that Ran-like may still interact with RanGAP, although potentially at a reduced level, as several residues within—and proximal to—known RanGAP interacting residues have suffered non-conservative mutations (e.g., amino acids 91-94, which are likely under positive selection, and amino acid 128 which is not suggested to be under positive selection) (Figure 2). Mapping these mutations onto co-crystal structures of Ran/RanGAP indicate that there are no major steric or charge conflicts. Further, RanGAP has been shown to produce duplicate genes (e.g., *Sd* (KUSANO *et al.* 2003) and to be under positive selection itself (PRESGRAVES 2007)), and interactions with a changing RanGAP or its duplicates could also explain the observed changes in Ran-like's RanGAP interface.

All other parental functional surfaces of Ran-like from the *D. melanogaster* subgroup seem to be even less conserved than the RanGAP interface. The analyses below focus on *D. melanogaster* Ran-like, but similar conclusions apply to the other Ran-like lineages analyzed. Regarding the interaction with Importin β, Ran-like proteins have likely reduced overall charge interactions through partial or complete changes in charge (E113G [disrupts hydrogen bond], K142T [in the basic patch], Y146L, and Y147I). In addition, one amino acid replacement (Q145Y) may possibly introduce a steric clash with position 163 (W163Y). The C-terminal end that is known to stabilize RanGDP (SEEWALD *et al.* 2002) may also have diverged. It is known that in the absence of this end, the RanGAP mediated hydrolysis of RanGTP to RanGDP is accelerated (SEEWALD *et al.* 2002), and the exchange of GDP to GTP catalyzed by RanGEF is also accelerated (RICHARDS *et al.* 1995). The C-terminal end is also required for the efficient binding of Ran to several of the Ran-binding proteins. Such binding is required for proper function of Ran (RICHARDS *et al.* 1995), but it is likely lost in Ran-like. RanGEF and exportins may also have a weaker interaction with Ran-like. Ran-like residues involved in RanGEF interaction have lost charge (partially or completely) or hydrophobicity (R95S, and V96N). Residue 95 is likely under positive selection (Table 2 and Figure 2). Ran-like residue 37 involved in exportin interaction has changed dramatically in charge and size (K37M).

The above analysis seems to indicate that the *D. melanogaster* subgroup *ran-like* protein has retained DNTF-2 and RanGAP interfaces while likely losing, or at least diminishing, all other protein-protein interfaces. The presence of DNTF-2 and RanGAP interfaces on Ran-like suggests that the Ran-like protein might exist in the RanGDP form and could be carried into the nucleus by DNTF-2r, where the Ran-like RanGDP could be converted to RanGTP. The Importin β interface on Ran-like, however, is likely diminished. As a result, Ran-GTP might not be transported out of the nucleus by Importin β, and Importin β might have a diminished capacity to release cargo upon nuclear entry. Similarly, export of RanGTP by the exportin complex may possibly be reduced. Additionally, the loss of Ran-like's C-terminal residues suggest that hydrolysis of RanGTP to RanGDP might possibly be accelerated in the presence of RanGAP. Other changes in Ran-like may affect the exchange of GDP to GTP by RanGEF. The binding of several lesser-known Ran binding proteins may also be affected. All these structural inferences remain to be experimentally tested. However, from these data it seems that Ran-like cannot completely replace Ran in testes in those *D. melanogaster* subgroup species where it is still functional.

References

BERMAN, H. M., T. BATTISTUZ, T. N. BHAT, W. F. BLUHM, P. E. BOURNE *et al.*, 2002 The Protein Data Bank. Acta Crystallogr D Biol Crystallogr **58:** 899-907.

BULLOCK, T. L., W. D. CLARKSON, H. M. KENT and M. STEWART, 1996 The 1.6 angstroms resolution crystal structure of nuclear transport factor 2 (NTF2). J Mol Biol **260:** 422-431.

ISGRO, T. A., and K. SCHULTEN, 2007 Association of nuclear pore FG-repeat domains to NTF2 import and export complexes. J Mol Biol **366:** 330-345.

KUSANO, A., C. STABER, H. Y. CHAN and B. GANETZKY, 2003 Closing the (Ran)GAP on segregation distortion in Drosophila. Bioessays **25:** 108-115.

KUSANO, A., C. STABER and B. GANETZKY, 2002 Segregation distortion induced by wild- type RanGAP in Drosophila. Proc Natl Acad Sci U S A **99:** 6866-6870.

MARCHLER-BAUER, A., J. B. ANDERSON, M. K. DERBYSHIRE, C. DEWEESE-SCOTT, N. R. GONZALES *et al.*, 2007 CDD: a conserved domain database for interactive domain family analysis. Nucleic Acids Res. **35:** D237-240.

MATSUURA, Y., and M. STEWART, 2004 Structural basis for the assembly of a nuclear export complex. Nature **432:** 872-877.

PRESGRAVES, D. C., 2007 Does genetic conflict drive rapid molecular evolution of nuclear transport genes in Drosophila? Bioessays **29:** 386-391.

QUIMBY, B. B., T. LAMITINA, S. W. L'HERNAULT and A. H. CORBETT, 2000 The mechanism of ran import into the nucleus by nuclear transport factor 2. J Biol Chem **275:** 28575-28582.

RENAULT, L., J. KUHLMANN, A. HENKEL and A. WITTINGHOFER, 2001 Structural basis for guanine nucleotide exchange on Ran by the regulator of chromosome condensation (RCC1). Cell **105:** 245-255.

RIBBECK, K., G. LIPOWSKY, H. M. KENT, M. STEWART and D. GORLICH, 1998 NTF2 mediates nuclear import of Ran. EMBO Journal **17:** 6587-6598.

RICHARDS, S., K. LOUNSBURY and I. MACARA, 1995 The C terminus of the nuclear RAN/TC4 GTPase stabilizes the GDP-bound state and mediates interactions with RCC1, RAN-GAP, and HTF9A/RANBP1. J Biol Chem **270:** 14405-14411.

SEEWALD, M. J., C. KORNER, A. WITTINGHOFER and I. R. VETTER, 2002 RanGAP mediates GTP hydrolysis without an arginine finger. Nature **415:** 662-666.

STEWART, M., H. M. KENT and A. J. MCCOY, 1998 Structural basis for molecular recognition between nuclear transport factor 2 (NTF2) and the GDP-bound form of the Ras-family GTPase Ran. J Mol Biol **277:** 635-646.

VETTER, I. R., A. ARNDT, U. KUTAY, D. GORLICH and A. WITTINGHOFER, 1999 Structural view of the Ran-Importin beta interaction at 2.3 A resolution. Cell **97:** 635-646.