

A third mitochondrial RNA polymerase in the moss *Physcomitrella patens*

Uwe Richter · Björn Richter · Andreas Weihe ·
Thomas Börner

Received: 10 April 2013/Revised: 16 July 2013/Accepted: 29 August 2013/Published online: 12 September 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract In most organisms, the mitochondrial genes are transcribed by RNA polymerases related to the single-subunit RNA polymerases of bacteriophages like T3 and T7. In flowering plants, duplication(s) of the *RpoTm* gene coding for the mitochondrial RNA polymerase (RPOTm) led to the evolution of additional RNA polymerases transcribing genes in plastids (RPOTp) or in both mitochondria and plastids (RPOTmp). Two putative RPOTmp enzymes were previously described to be encoded by the nuclear genes *RpoTmp1* and *RpoTmp2* in the moss *Physcomitrella patens*. Here, we report on a third *Physcomitrella RpoT* gene. We determined the sequence of the cDNA. Comparison of the deduced amino acid sequence with sequences of plant organellar RNA polymerases suggests that this gene encodes a functional phage-type RNA polymerase. The 78 N-terminal amino acids of the putative RNA polymerase were fused to GFP and found to target the fusion protein exclusively to mitochondria in *Arabidopsis* protoplasts. *P. patens* is the only known organism to possess three mitochondrial RNA polymerases.

Keywords Chloroplast transcription · Mitochondrial transcription · NEP · PEP · Phage-type RNA polymerase · *Physcomitrella*

Introduction

Mitochondrial transcription is strictly dependent on nuclear gene-encoded phage-type RNA polymerases since all eukaryotes, with the exception of the jacobid protists, seem to lack a mitochondrial bacterial-type RNA polymerase (Burger et al. 2013). On the contrary, chloroplasts have retained a eubacterial RNA polymerase with chloroplast gene-encoded core subunits and nuclear gene-encoded sigma factor(s) from their cyanobacterial ancestor. In angiosperms, the bacterial-type RNA polymerase (called PEP for plastid-encoded plastid RNA polymerase) is not sufficient for transcription of the chloroplast genes. The activity of one or two phage-type RNA polymerase(s) (called NEP for nuclear-encoded plastid RNA polymerase) has to complement PEP activity. NEP has evolved in angiosperms by duplication(s) of the *RpoTm* gene for the mitochondrial RNA polymerase (*RpoT* for RNA polymerase of phage T3/7-type; *m* for mitochondrial) (Liere et al. 2011). *RpoT* genes encoding mitochondrial and/or plastid phage-type polymerases have been identified in several land plants. Eudicotyledonous plants such as *Arabidopsis* and *Nicotiana* possess at least three *RpoT* genes coding for a mitochondrial (RPOTm), a plastid (RPOTp) and a dual-targeted polymerase (RPOTmp) (Hedtke et al. 2000, 2002; Kobayashi et al. 2002, a, b). Nuclear genomes of monocots investigated thus far harbour not more than two *RpoT* genes, one coding for a mitochondrial RNA polymerase (RPOTm) and the other for a plastidial enzyme (RPOTp) (Chang et al. 1999; Emanuel et al. 2004; Ikeda

Communicated by R. Bock.

Electronic supplementary material The online version of this article (doi:10.1007/s00294-013-0405-y) contains supplementary material, which is available to authorized users.

U. Richter · B. Richter · A. Weihe · T. Börner (✉)
Institut für Biologie-Genetik, Humboldt-Universität zu Berlin,
Chausseestr. 117, 10115 Berlin, Germany
e-mail: thomas.boerner@rz.hu-berlin.de

Present Address:

U. Richter
Research Programs Unit, Molecular Neurology, University of
Helsinki, Haartmaninkatu 8, 00290 Helsinki, Finland

AtRpoTm M-----W--RN ILGRA LRVKVK FL-----SDSSSSG THY
 AtRpoTm M--S-----Q TPLFLANQ TKV FDDLILPHKFP ISSPNVSO SFPWNRN I
 AtRpoTm MA SA-----A--AA SPSL SLNPTSHFQHQ TSLV TW--LK PPPSALFRRK TLP
 PpRpoT3 M-----W--RAAARY FK SEL LHCGG SR SR IAQV DLYA--LA QHRL LCSATA
 PpRpoTm1 MVA IGVLEPISAVGR TGRRI SDVKLVGLRSHQDCAPTFNLGGRVRRGMWRAAVRQLSRQ PR EGR LGAGNCSSLFW SQ SLQSRW TSG SAAAA
 PpRpoTm2 MVA IGVLEPISAVGR TGRRI SDVKLVGLRSHQDCAPTFNLGGRVRRGMWRAAVRQLSRQ PR EGR LGAGNCSSLFW SQ SLQSRW TSG SAAAA

AtRpoTm PVNRVRG ILSVNL SGVRNGL-----SINPVNE-----MG-----GSSFRHGQC VVEGYATAQA IDS TDPDED
 AtRpoTm AKQAI SR SAAR LNV SSQ TRGLL-----VSPSEI FSKNLSFR FVVLG SPCHGKGR CLSGH TRREEFPSK SER CLSGT LARGY TSVAE EEEV
 AtRpoTm FFER HSLPISA SSSSSS-----SSSTLSV HEK PISNSVH-----FHGN IESFENQDSSYAGTIKGA SLIELEEN PV
 AtRpoTm EVQTVDR LCGSSSDHIF-----SPSSLSLQ LNTPLEDILD-----AALLGV SRYDNRDFR SLN FYR SAV FSS TAEPA
 PpRpoT3 VGVVHVR AVDK DVNV DARW FLRDTKVG FSRSSSD SVPSSTLDQ TLDAALHN VVNLPW LFLQGV PAQ DVGNF HAR PGER PALDY CRTYASAAE
 PpRpoTm1 SVQSLDRA SE DR S SEFK SGLR-----FLK P SSS SPS SLT PLDSA FIELVDLLPKCQWSTTTDFN HDH SKLQ TFLQV FDSRAYSAL
 PpRpoTm2 SVQSLDRA SE DR S SEFK SGLR-----FLK P SSS SPS SLT PLDSA FIELVDLLPKCQWSTTTDFN HDH SKLQ TFLQV FDSRAYSAL

AtRpoTm SSGSDV NELITM--EKETER IRKKA LAA IPPRV IAGMGAQ KPYM LKROWK M ETE WBRAR BCRBELA DMCOR LAPNLEYM RSTF
 AtRpoTm LSTDV EEPVVD LKEMK EKKR ESHR SWRMK-KDDQ FGMGR R FQNL RROWK HETE WBRAR AAE MBIEL DMCOR LAPNLEYM RSTF
 AtRpoTm RNGLSGRRRLFM DDPWISALFLKGLKSMV DQTI KIERK D IDKR FDS LKROWK HETAW R MVSD R D R E M C K LAPNLEYM RSTF
 AtRpoTm YIGYD PPEGTEV DYDERG IEFMSEHAGAAVA KEMERKKGN LQ L R R O I N E T B A W I N A E O M E T E M C K LAPNLEYM RSTF
 PpRpoT3 AVIDDDES--E ECAPE PPO SVDFN KAK TAK EAAE SRKARW EKA L R R O K E T B A W O A A E K E L V M C M C K LAPNLEYM RSTF
 PpRpoTm1 EDELEDEDI GK ENEHAR EALPNG SNWDTGLDASLEOK SRKR RAR F L R R O W K H E T B A W O A A T E M R E L M M C M C K LAPNLEYM RSTF
 PpRpoTm2 EDELEDEDI GK ENEHAR EALPNG SNWDTGLDASLEOK SRKR RAR F L R R O W K H E T B A W O A A T E M R E L M M C M C K LAPNLEYM RSTF

AtRpoTm LWFEPV R N A T O P L D T E--K I R K G K T E V A P M D O L P A D K M A V I T M H R M G I L M N A B S V G V K L V N A A T G C E A V E Q E V R I N S F Q K R N
 AtRpoTm LWFEPV R N A T O P L D T E--K I R K G K T E V A P M D O L P A D K M A V I T M H R M G I L M N A B S V G V K L V N A A T G C E A V E Q E V R I N S F Q K R N
 AtRpoTm LWFEPV R N A T O P L D T E--K I R K G K T E V A P M D O L P A D K M A V I T M H R M G I L M N A B S V G V K L V N A A T G C E A V E Q E V R I N S F Q K R N
 PpRpoT3 LWFEPV R N A T O P L D T E--K I R K G K T E V A P M D O L P A D K M A V I T M H R M G I L M N A B S V G V K L V N A A T G C E A V E Q E V R I N S F Q K R N
 PpRpoTm1 LWFEPV R N A T O P L D T E--K I R K G K T E V A P M D O L P A D K M A V I T M H R M G I L M N A B S V G V K L V N A A T G C E A V E Q E V R I N S F Q K R N
 PpRpoTm2 LWFEPV R N A T O P L D T E--K I R K G K T E V A P M D O L P A D K M A V I T M H R M G I L M N A B S V G V K L V N A A T G C E A V E Q E V R I N S F Q K R N

AtRpoTm K--KNA TDKTINTE--ANVSEI VAKETE K-----ARQVTV LHEKN KLRQVKA LVRK HDSFR PEGQEAQV KVGARLIQLM ENAV
 AtRpoTm K-----GDDNEE--SGGVEN ETSKKEODK-----LRKVN E L R K O K L S A V R K T L O S H D Y T N P W I A D V R A K V G R L L E L L V R T A V Y
 AtRpoTm K-----NN--AGDSQ E L K E Q L-----LRKVN E L R K O K L S A V R K T L O S H D Y T N P W I A D V R A K V G R L L E L L V R T A V Y
 AtRpoTm K-KNETKRKG TKD--DGIST I E S G V E L T T S P E L--P Q A T G M K L R K L R L W R S L L O K A S G S D P W C T I Y A K V G R L L E L L V R T A V Y
 PpRpoT3 K V A K V K N K V S G D A L D L D L D T N D D S I A N S V L D S D L K L L A K V K K L W K O K L R L V G K V V O Q A N D E P W G P A Q V K V G R L L E L L V R T A V Y
 PpRpoTm1 KSKDAE--GN TNSLGDVANNEEL EATRKA D L K S K V--V R K V K L W K O K L R L V D S L I K H A S T D P W S T V I H K V G R L L E L L V R T A V Y
 PpRpoTm2 KSKDAE--GN TNSLGDVANNEEL EATRKA D L K S K V--V R K V K L W K O K L R L V D S L I K H A S T D P W S T V I H K V G R L L E L L V R T A V Y

AtRpoTm O P A B O P D D G P P D I R P A F K O N F R V T L - E N K T S R R G C T E C D P V L K G L D K S R H M V L P Y L P M L H Q N Q M C M D O G H H E L S V M R T H G
 AtRpoTm O P A B O P D D G P P D I R P A F K O N F R V T L - E N K T S R R G C T E C D P V L K G L D K S R H M V L P Y L P M L H Q N Q M C M D O G H H E L S V M R T H G
 AtRpoTm O P A B O P D D G P P D I R P A F K O N F R V T L - E N K T S R R G C T E C D P V L K G L D K S R H M V L P Y L P M L H Q N Q M C M D O G H H E L S V M R T H G
 PpRpoT3 O P A B O P D D G P P D I R P A F K O N F R V T L - E N K T S R R G C T E C D P V L K G L D K S R H M V L P Y L P M L H Q N Q M C M D O G H H E L S V M R T H G
 PpRpoTm1 R P A B O P D D G P P D I R P A F K O N F R V T L - E N K T S R R G C T E C D P V L K G L D K S R H M V L P Y L P M L H Q N Q M C M D O G H H E L S V M R T H G
 PpRpoTm2 R P A B O P D D G P P D I R P A F K O N F R V T L - E N K T S R R G C T E C D P V L K G L D K S R H M V L P Y L P M L H Q N Q M C M D O G H H E L S V M R T H G

AtRpoTm AKQOQV T K R T F K O I E F V E A L D L G N K K K I K K V L S L V D R I N S G G R I S C L V D R E D V F I P E P P E R E D E K K K N R W E S K K A K O N N E R
 AtRpoTm AKQOQV T K R T F K O I E F V E A L D L G N K K K I K K V L S L V D R I N S G G R I S C L V D R E D V F I P E P P E R E D E K K K N R W E S K K A K O N N E R
 AtRpoTm SKKQD A K D I S H T A H R V F E A L D L G N K K R V R N K V L V V E R L W A G G R F A C L V N R E D V F I P E P P E R E D E K K K N R W E S K K A K O N N E R
 AtRpoTm SK E O V D I F K N T P R E N K K I F O A L N L G T O W R V N K V L V L E O I W K G G R F A N L V D A E D V I P A P E N N D B E L K S W R E V G I V R N R N E R
 PpRpoT3 A K E I D E T I S L R O D M I K I O A L D L G T O W K I N N V L V L E O M W K G G R F A N L V D A E D V I P A P E N N D B E L K S W R E V G I V R N R N E R
 PpRpoTm1 A K E O R M A V I D T P R K H M K V V E A L N L G T O W R V N K V L V V K L W K G G R F A N L V D A E D V I P A P E N N D B E L K S W R E V G I V R N R N E R
 PpRpoTm2 A K E O R M A V I D T P R K H M K V V E A L N L G T O W R V N K V L V V K L W K G G R F A N L V D A E D V I P A P E N N D B E L K S W R E V G I V R N R N E R

Block III

AtRpoTm HSRCDF ELK T F M A R K M D E E F Y P H N V D F R C R A Y P M F L N H L G S D L C R G I L F A D G R K L G S G L W L K I H A N Y I G V K M S F D R I
 AtRpoTm HSRCDF ELK T F M A R K M D E E F Y P H N M D F R C R A Y P M F L N H L G S D L C R G I L F A D G R K L G S G L W L K I H A N Y I G V K M S F D R I
 AtRpoTm HSRCDF ELK T F M A R K M D E E F Y P H N M D F R C R A Y P M F L N H L G S D L C R G I L F A D G R K L G S G L W L K I H A N Y I G V K M S F D R I
 PpRpoT3 Y S R C D F E L K T A M A R K T N E D F Y P H N L D F R C R A Y P M F L N H L G S D M C R G I L F A D G R K L G S G L W L K I H A N Y I G S K S S F D R I
 PpRpoTm1 H S V R C D F E L K T A M A R K T D E E F F Y P H N L D F R C R A Y P M F L N H L G S D M C R G I L F A D G R K L G S G L W L K I H A N Y I G S K S S F D R I
 PpRpoTm2 H S V R C D F E L K T M A K K T D E E F Y P H N L D F R C R A Y P M F L N H L G N D L C R G I L F A D G R K L G S G L W L K I H A N Y I G S V K M S F D R I

Block IV

AtRpoTm A E T E H L D I F D S S R P L R G R M W I A E D P P O C L A A C T I N S R A I N S P I F E A A I S H I P I H O D G S C N G L O H Y A A L G R D I G A D V N L V A G E R P
 AtRpoTm A E T E H L D I F D S A R P L R G R M W I A E D P P O C L A A C T I N S R A I N S P I F E A A I S H I P I H O D G S C N G L O H Y A A L G R D I G A D V N L V A G E R P
 AtRpoTm A E T E H L D I F D S A R P L R G R M W I A E D P P O C L A A C T I N S R A I N S P I F E A A I S H I P I H O D G S C N G L O H Y A A L G R D I G A D V N L V A G E R P
 AtRpoTm A H V D H M D D V F D S A R P L R G R M W I A E D P P O C L A A C T I N S R A I N S P I F E A A I S H I P I H O D G S C N G L O H Y A A L G R D I G A D V N L V A G E R P
 PpRpoT3 A M V D E V M E K V F D S A R P L R G R M W I A E D P P O C L A A C T I N S R A I N S P I F E A A I S H I P I H O D G S C N G L O H Y A A L G R D I G A D V N L V A G E R P
 PpRpoTm1 A M A E D R M E E I F D S A R P L R G R M W I A E D P P O C L A A C T I N S R A I N S P I F E A A I S H I P I H O D G S C N G L O H Y A A L G R D I G A D V N L V A G E R P
 PpRpoTm2 A M A E D R M E E I F D S A R P L R G R M W I A E D P P O C L A A C T I N S R A I N S P I F E A A I S H I P I H O D G S C N G L O H Y A A L G R D I G A D V N L V A G E R P

Block V

AtRpoTm A D V Y T I A R V K T M O O D A E D P E T F P N A T V A K M M L O V D R K L V K O T V M T S V Y G V T F V G A R O I K R R L K E R G T F E D S L T E H A S C Y A A K I
 AtRpoTm A D V Y T I A R V K T M O O D A E D P E T F P N A T V A K M M L O V D R K L V K O T V M T S V Y G V T F V G A R O I K R R L K E R G T F E D S L T E H A S C Y A A K I
 AtRpoTm A D V Y T I A R V K T M O O D A E D P E T F P N A T V A K M M L O V D R K L V K O T V M T S V Y G V T F V G A R O I K R R L K E R G T F E D S L T E H A S C Y A A K I
 AtRpoTm A D V Y S I T A R V K T M O O D A E D P E T F P N A T V A K M M L O V D R K L V K O T V M T S V Y G V T F V G A R O I K R R L K E R G T F E D S L T E H A S C Y A A K I
 PpRpoT3 A D V Y S I T A R V K T M O O D A E D P E T F P N A T V A K M M L O V D R K L V K O T V M T S V Y G V T F V G A R O I K R R L K E R G T F E D S L T E H A S C Y A A K I
 PpRpoTm1 A D V Y S I T A R V K T M O O D A E D P E T F P N A T V A K M M L O V D R K L V K O T V M T S V Y G V T F V G A R O I K R R L K E R G T F E D S L T E H A S C Y A A K I
 PpRpoTm2 A D V Y S I T A R V K T M O O D A E D P E T F P N A T V A K M M L O V D R K L V K O T V M T S V Y G V T F V G A R O I K R R L K E R G T F E D S L T E H A S C Y A A K I

Block VI

AtRpoTm T K A L F E M F E A R I K S W F S D C A K I A S E N N A V Q W T P L G L P V O P Y R K P G R H L V K T L O V L H L S R--E T D K V M A R R O N T A E P N F I H S L D
 AtRpoTm T K A L F E M F E A R I K S W F S D C A K I A S E N N A V Q W T P L G L P V O P Y R K P G R H L V K T L O V L H L S R--E T D K V M A R R O N T A E P N F I H S L D
 AtRpoTm T K A L F E M F E A R I K S W F S D C A K I A S E N N A V Q W T P L G L P V O P Y R K P G R H L V K T L O V L H L S R--E T D K V M A R R O N T A E P N F I H S L D
 AtRpoTm T K A L F E M F E A R I K S W F S D C A K I A S E N N A V Q W T P L G L P V O P Y R K P G R H L V K T L O V L H L S R--E T D K V M A R R O N T A E P N F I H S L D
 PpRpoT3 T K A L F E M F E A R I K S W F S D C A K I A S E N N A V Q W T P L G L P V O P Y R K P G R H L V K T L O V L H L S R--E T D K V M A R R O N T A E P N F I H S L D
 PpRpoTm1 T K A L F E M F E A R I K S W F S D C A K I A S E N N A V Q W T P L G L P V O P Y R K P G R H L V K T L O V L H L S R--E T D K V M A R R O N T A E P N F I H S L D
 PpRpoTm2 T K A L F E M F E A R I K S W F S D C A K I A S E N N A V Q W T P L G L P V O P Y R K P G R H L V K T L O V L H L S R--E T D K V M A R R O N T A E P N F I H S L D

Block VII

AtRpoTm G S H M M T A V A C N R A G L S F A G V H D S F W T H A C D V D M N L I R E K F V E L Y E K P I L E N L L E S F O K S F P P I S F P P L P R G D L D R K V L E S Y F F N
 AtRpoTm G S H M M T A V A C N R A G L S F A G V H D S F W T H A C D V D M N L I R E K F V E L Y E K P I L E N L L E S F O K S F P P I S F P P L P R G D L D R K V L E S Y F F N
 AtRpoTm G S H M M T A V A C N R A G L S F A G V H D S F W T H A C D V D M N L I R E K F V E L Y E K P I L E N L L E S F O K S F P P I S F P P L P R G D L D R K V L E S Y F F N
 AtRpoTm G S H M M T A V A C N R A G L S F A G V H D S F W T H A C D V D M N L I R E K F V E L Y E K P I L E N L L E S F O K S F P P I S F P P L P R G D L D R K V L E S Y F F N
 PpRpoT3 G S H M M T A V A C N R A G L S F A G V H D S F W T H A C D V D M N L I R E K F V E L Y E K P I L E N L L E S F O K S F P P I S F P P L P R G D L D R K V L E S Y F F N
 PpRpoTm1 G S H M M T A V A C N R A G L S F A G V H D S F W T H A C D V D M N L I R E K F V E L Y E K P I L E N L L E S F O K S F P P I S F P P L P R G D L D R K V L E S Y F F N
 PpRpoTm2 G S H M M T A V A C N R A G L S F A G V H D S F W T H A C D V D M N L I R E K F V E L Y E K P I L E N L L E S F O K S F P P I S F P P L P R G D L D R K V L E S Y F F N

Block VIII

AtRpoTm S H M M T A V A C N R A G L S F A G V H D S F W T H A C D V D M N L I R E K F V E L Y E K P I L E N L L E S F O K S F P P I S F P P L P R G D L D R K V L E S Y F F N
 AtRpoTm S H M M T A V A C N R A G L S F A G V H D S F W T H A C D V D M N L I R E K F V E L Y E K P I L E N L L E S F O K S F P P I S F P P L P R G D L D R K V L E S Y F F N
 AtRpoTm S H M M T A V A C N R A G L S F A G V H D S F W T H A C D V D M N L I R E K F V E L Y E K P I L E N L L E S F O K S F P P I S F P P L P R G D L D R K V L E S Y F F N
 AtRpoTm S H M M T A V A C N R A G L S F A G V H D S F W T H A C D V D M N L I R E K F V E L Y E K P I L E N L L E S F O K S F P P I S F P P L P R G D L D R K V L E S Y F F N
 PpRpoT3 S H M M T A V A C N R A G L S F A G V H D S F W T H A C D V D M N L I R E K F V E L Y E K P I L E N L L E S F O K S F P P I S F P P L P R G D L D R K V L E S Y F F N
 PpRpoTm1 S H M M T A V A C N R A G L S F A G V H D S F W T H A C D V D M N L I R E K F V E L Y E K P I L E N L L E S F O K S F P P I S F P P L P R G D L D R K V L E S Y F F N
 PpRpoTm2 S H M M T A V A C N R A G L S F A G V H D S F W T H A C D V D M N L I R E K F V E L Y E K P I L E N L L E S F O K S F P P I S F P P L P R G D L D R K V L E S Y F F N

◀ **Fig. 1** Amino acid sequence alignment. Amino acid sequences were compared among *Arabidopsis* and *Physcomitrella* RPOT polymerases using the ClustalW algorithm. Sequences with accession numbers: AtRPOTm (CAA69331), AtRPOTp (CAA70210), AtRPOTmp (CAC17120); PpRPOTmp1 (CAC95163); PpRPOTmp2 (CAC95164). Grey boxes indicate conserved blocks in the RPOT polymerase family; functionally crucial residues (McAllister and Raskin 1993; Sousa et al. 1993) are indicated by asterisks

and Gray 1999), seemingly lacking the *RpoTmp* gene encoding the dual-targeted RPOT polymerase found in eudicots. This *RpoTmp* originated most likely after *RpoTp* from a second duplication of the *RpoTm* gene during angiosperm evolution (Liere et al. 2011). In the basal angiosperm *Nuphar advena*, two duplications of the *RpoTm* gene gave rise to the plastid enzyme RPOTp and to a second mitochondrial RNA polymerase (Yin et al. 2010).

While *RpoT* genes from angiosperms have been well characterised [reviewed in (Liere et al. 2011)], almost nothing is known about the evolution of those genes in other lineages of the plant kingdom. Only one *RpoT* gene coding for the mitochondrial RNA polymerase was detected in the lycophyte *Selaginella moellendorffii* (Yin et al. 2009). In the deeper branching leafy moss *Physcomitrella patens*, however, two nuclear *RpoT* genes were reported to exist (Kabeya et al. 2002; Richter et al. 2002). Dual targeting of the two *RpoT* gene products to both mitochondria and plastids being the result of two different translational starts has been described, and the polymerases were designated as PpRPOTmp1 and PpRPOTmp2 (Richter et al. 2002). For both PpRPOTmp1 and PpRPOTmp2, translation initiation at the first of two in-frame AUG start codons was found to compose a polypeptide with N-terminal plastidial targeting properties, whereas initiation at the downstream AUG gave rise to a mitochondrial protein (Richter et al. 2002). More recently, Kabeya and Sato (2005) demonstrated that translation initiation from the first AUG at the *RpoTmp1* and *RpoTmp2* transcripts was repressed for all constructs with the native 5'UTR present. Thus, targeting to mitochondria or chloroplasts of these enzymes might be tightly regulated by alternative usage of the two start codons.

Phylogenetic analyses show a distinct group of the plant *RpoT* sequences clearly separated from those of fungi and animals. In concordance with this finding, the plant specific gene structure is highly conserved in all plant *RpoT* genes investigated thus far (Richter et al. 2002). Within this group, the RPOT polymerases of *Physcomitrella patens* are separated from the angiosperm polymerases, with the latter falling into two groups with plastid-localised enzymes on one hand, and mitochondrial- and dual-targeted polymerases on the other (Liere and Börner 2007; Richter et al. 2002). The duplication event that led to the *RpoT* gene

copies in *P. patens*, thus, must be considered to have occurred independently from the one giving rise to the *RpoT* gene family in angiosperms. Furthermore, it shows that the duplication giving rise to a group of exclusively plastid-localised RPOTp gene products of angiosperms is predating the duplication event that led to the appearance of the third gene as found in eudicot plants, such as *A. thaliana* or *N. sylvestris* (Richter et al. 2002).

The moss *P. patens* (Hedw.) Bruch and Schimp is a widely spread species which colonizes open habitats in cold temperate zones. The occurrence of homologous recombination in its nuclear genome and several other aspects of its biology, e.g., small size, short life cycle, a relatively easy cultivation procedure, its basal phylogenetic position, an extensive EST database and the fact that it is predominantly haploid make *Physcomitrella* ideal for molecular and comparative studies (Cove 2005). Moreover, the complete genome sequence has been published (Rensing et al. 2008).

Surprisingly, BLAST analysis of the genome sequence of *P. patens* revealed a third genomic locus (provisionally designated as *PpRpoT3*) bearing close similarity to plant phage-type RNA polymerases. We have determined the sequence of this third *Physcomitrella RpoT* gene and of its transcript. We report here that it encodes an RNA polymerase that is targeted to the mitochondria.

Materials and methods

PpRpoT3 cDNA cloning

The BLAST analysis of the published *Physcomitrella* whole genome sequence (Rensing et al. 2008) revealed a third genomic locus potentially encoding a phage-type RNA polymerase on PHYPA scaffold_241 (NW_001865497). Putative coding regions were used to generate primers for the amplification of cDNA ends. 3'- and 5'-RACE reactions employed the CapFishing Full-length cDNA Premix kit (Seegene; Eschborn, Germany) and Phusion hot start DNA polymerase (New England Biolabs GmbH, Frankfurt am Main, Germany) according to the protocols of the manufacturers. For cDNA synthesis, 1 µg of polyA-enriched *Physcomitrella* protonema RNA was reverse transcribed using SuperScript™ III RNase H-reverse transcriptase kit (Life Technologies, Darmstadt, Germany). Notably, 5'RACE reverse transcription employed two gene-specific primers annealing to the in silico-predicted lower exon 1 region (5'acePpT3gsp1/5'acePpT3gsp2). While 3'- and 5'-RACE adaptor primers were provided in the kit, gene-specific primers (gsp) were employed for two rounds of PCR to enhance the specificity of the reaction (see Supplementary Table 1). To verify the

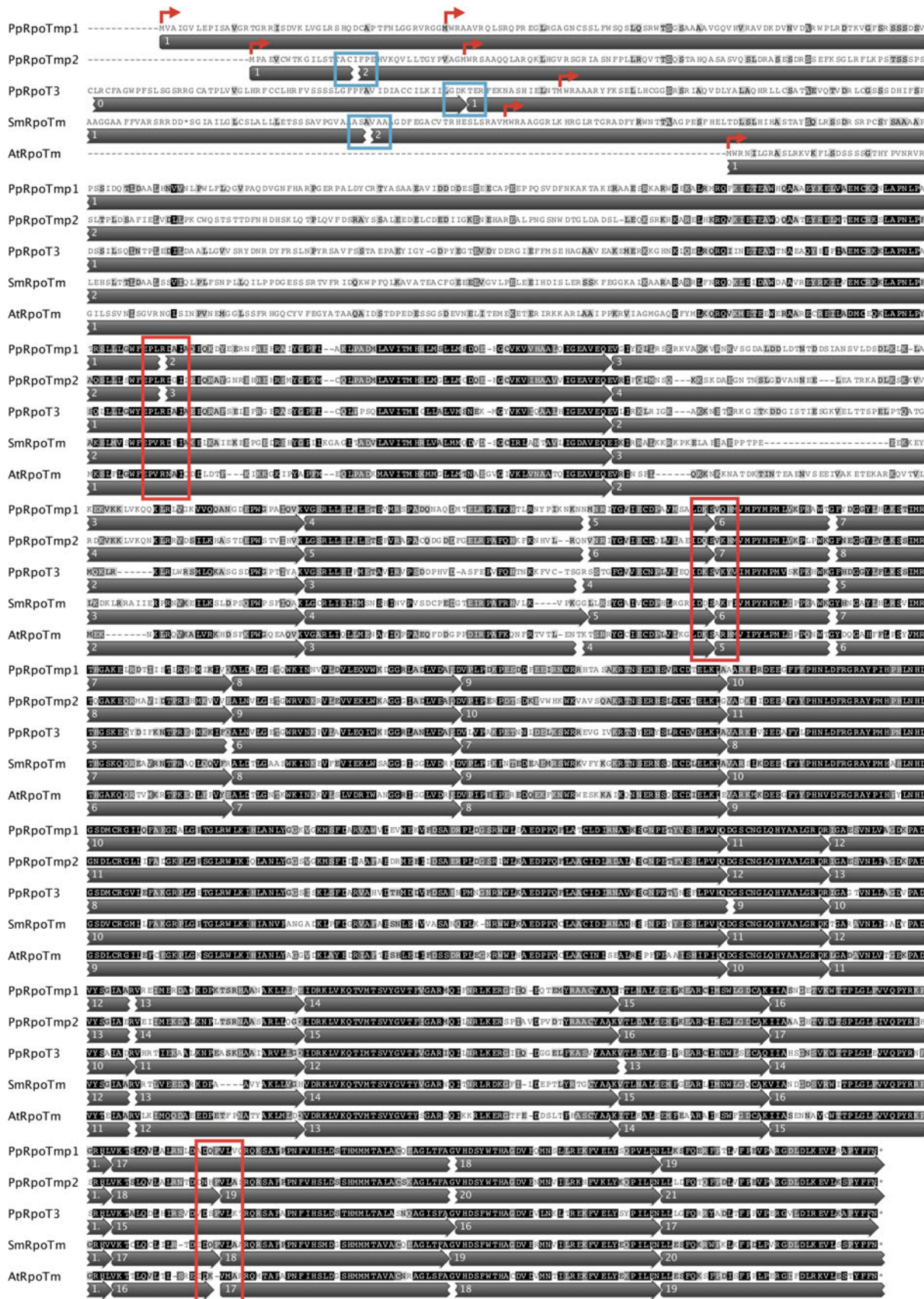


Fig. 2 ClustalW-derived protein alignment. All three RPOT proteins from *Physcomitrella* were aligned with SmRPOTm of *Selaginella moellendorffii* and AtRPOTm of *Arabidopsis thaliana* using the ClustalW algorithm. Grey arrows below the sequence indicate exons and numbers inside indicate exon numbers. While red arrows mark putative translational start sites, blue boxes highlight intronic insertions in the putative 5'UTRs. Red boxes mark intron positions found in at least one of the *Physcomitrella RpoT* genes, but lacking in *PpRpoT3*

cloned transcript ends, a primer walking strategy was used until no specific amplification products were obtained after a second round of 35 PCR cycles (5racePpT3gsp7 and 3racePpT3gsp4). After the identification of both *PpRpoT3* transcript termini, the whole *PpRpoT3* coding region (the complete cDNA sequence was deposited in the EMBL database under accession number Hx2000029983) was amplified using primers PpT3fw1 and PpT3m3169, and the fragment was cloned into pDrive (pDriveRpoT3cDNA). To map the *PpRpoT3* gene structure, subsequent sequencing of the whole *PpRpoT3* cDNA employed primers used in previous steps of the cloning procedure. Sequencing primers are shown in Supplementary Table 1.

Generation of PpRPOT3-GFP fusion constructs

To generate the pPpRPOT3-GFP constructs and to drive the expression of the fusion protein PpRPOT3-GFP, a *PpRpoT3* fragment containing the full 5'UTR, and encoding 78 N-terminal amino acids of PpRPOT3 was amplified from pDriveRpoT3cDNA using primers PpT3gfpfwXbaI (cactctagaCTGCTTGCGTTGCTTTGC) and PpT3gfpPrevSalI (atagtcgacCAAGGATGTCTTCCAGAGGTG; lower-case letters correspond to non-annealing nucleotides for the introduction of restriction sites). PpT3gfpXbaI-2 (cactctagaGAGTTGAATACTATGTGGC) was used for the amplification of a construct lacking all, but 12 nt of the 5'UTR. PCR products were *XbaI/SalI*-digested and ligated into the *SpeI/SalI*-cleaved vector pOL-GFP S65C (Peeters et al. 2000). Mitochondrial CoxIV-GFP and plastidial RecA-GFP (Peeters et al. 2000) control constructs were kindly provided by Ian Small (UWA, Perth, Australia).

Transient expression in Arabidopsis protoplasts and microscopy

Protoplasts were isolated from *Arabidopsis thaliana* leaves grown for 4 weeks under long day conditions (23 °C, 16/8 h light/dark) as described before (Yoo et al. 2007). Briefly, 100 µl protoplasts (2×10^6 protoplasts/ml) were transfected with 20 µg pOL-PpRpoT3gfp plasmid in 40 % polyethylene glycol 4000, 0.8 M mannitol, and 1 mM CaCl₂. Transformed protoplasts were kept in the dark and

examined 48 h after transfection by confocal laser scanning microscopy with a Leica TCS SP2 using 488-nm excitation and two-channel measurement of emission from 510 to 580 nm (green/GFP) and >590 nm (red/chlorophyll).

Phylogenetic analyses

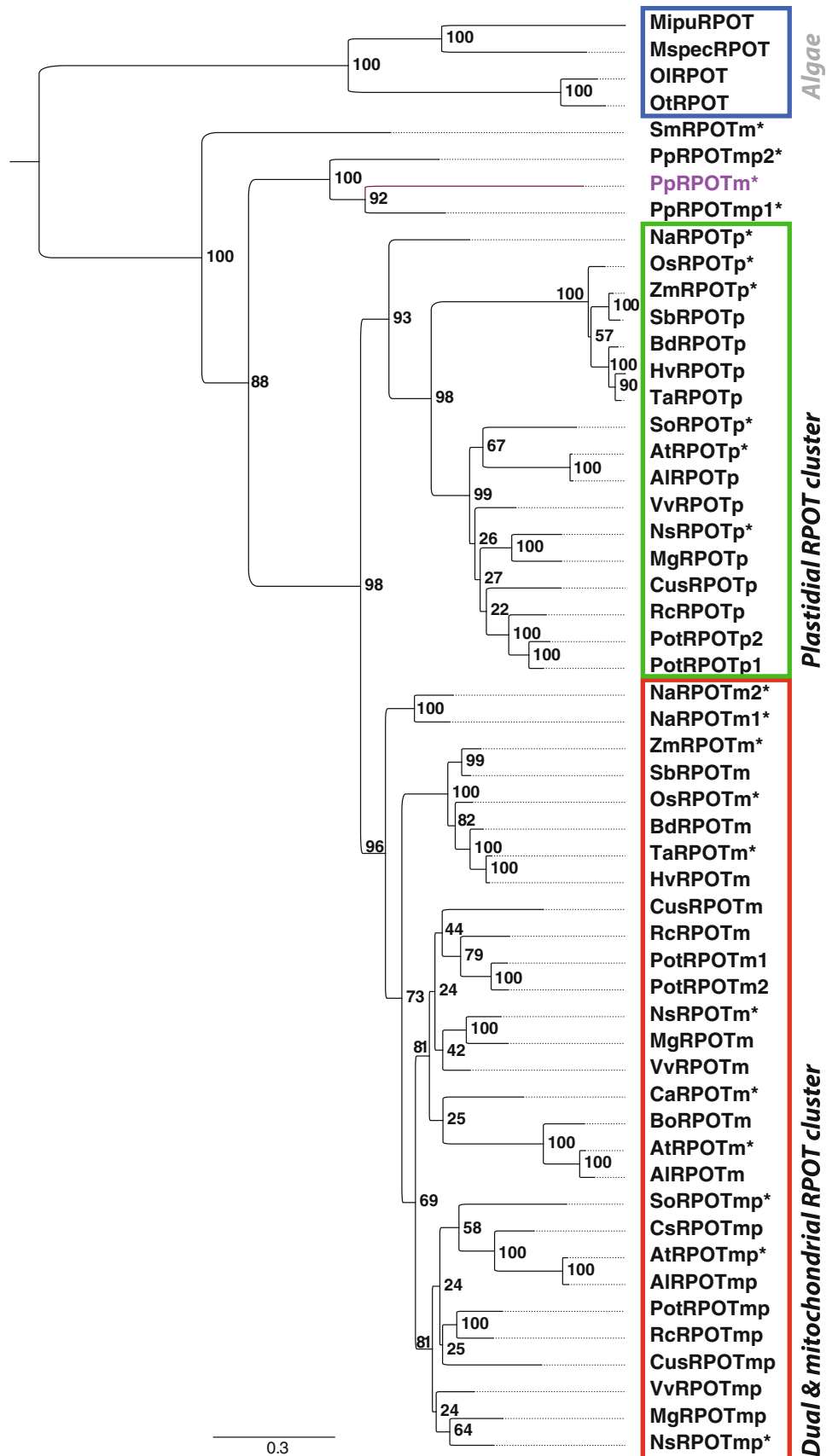
Sequence information was retrieved from the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/BLAST>) employing the BLASTP and TBLASTN algorithms and from the Joint Genome Institute (<http://genome.jgi-psf.org/>). For translation and alignment, sequences were subsequently imported into GENEIOUS (Drummond et al. 2010). Multiple protein sequence alignments were generated using CLUSTALW (Thompson et al. 1994) implemented in the GENEIOUS package. The phylogeny was reconstructed by the PhyML algorithm V3.0 (Guindon et al. 2010) with 100 bootstrap replicates. The aligned sequence data comprised 894 amino acids lacking only the less conserved N-terminal portion of the proteins. For the reconstruction, the LG amino acid replacement model was employed. A discrete gamma distribution with four categories was assumed to approximate the continuous function (shape: 1,004/inv: 0,138). Essentially, the same tree topology was obtained employing Bayesian (Ronquist and Huelsenbeck 2003) and neighbor joining analysis (Saitou and Nei 1987). The tree was rooted with the four green algae RPOT proteins indicated. Tree reconstruction was based on a multiple alignment of 54 plant RPOT sequences (Supplementary Table 2).

Results and discussion

Characterization of the third gene encoding a phage-type RNA polymerase in *Physcomitrella*

To identify the precise position and length of the open reading frame encoding the third phage-type RNA polymerase, we determined the full-length cDNA sequence. Transcripts of the gene were present in RNA preparations from protonema tissue, i.e., the *RpoT3* gene is active. The full-length cDNA comprising 3,496 bp was obtained by aligning fragments produced by RT-PCR and RACE. It constitutes 235 bp of untranslated leader, the protein coding sequence of 3,030 bp, and a 3'-untranslated sequence of 231 bp. Therefore, the predicted PpRPOT3 protein comprises 1,010 amino acids and it shows 46–48 % identity with the three *Arabidopsis* RPOTs. While the two *Physcomitrella* RPOT polypeptides RPOTmp1 and RPOTmp2 share 59 % identity, the scores for RPOT3 are significantly lower; 53 % with RPOTmp1 and 51 % with

Fig. 3 Phylogenetic analysis of plant RPOT proteins. The phylogeny was reconstructed by the PhyML algorithm V3.0 (Guindon et al. 2010) based on a multiple alignment of 54 plant RPOT sequences (listed in Supplementary Table 2) with 100 bootstrap replicates. Prefixes of designations of plant RpoT proteins refer to the abbreviations used in Supplementary Table 2. The tree was rooted with four *green* algae proteins. Asterisks indicate that experimental evidence for the localization in mitochondria, plastids or both organelles exists. Otherwise, the indicated localization (RPOTm-mitochondrial enzyme, p-plastid enzyme, and mp-enzyme targeted to both mitochondria and plastids) is solely based on the position in the tree. Based on the experimental data shown in Fig. 4, *PpRpoT3*/*PpRPOT3* was renamed into *PpRpoTm*/*PpRPOTm*



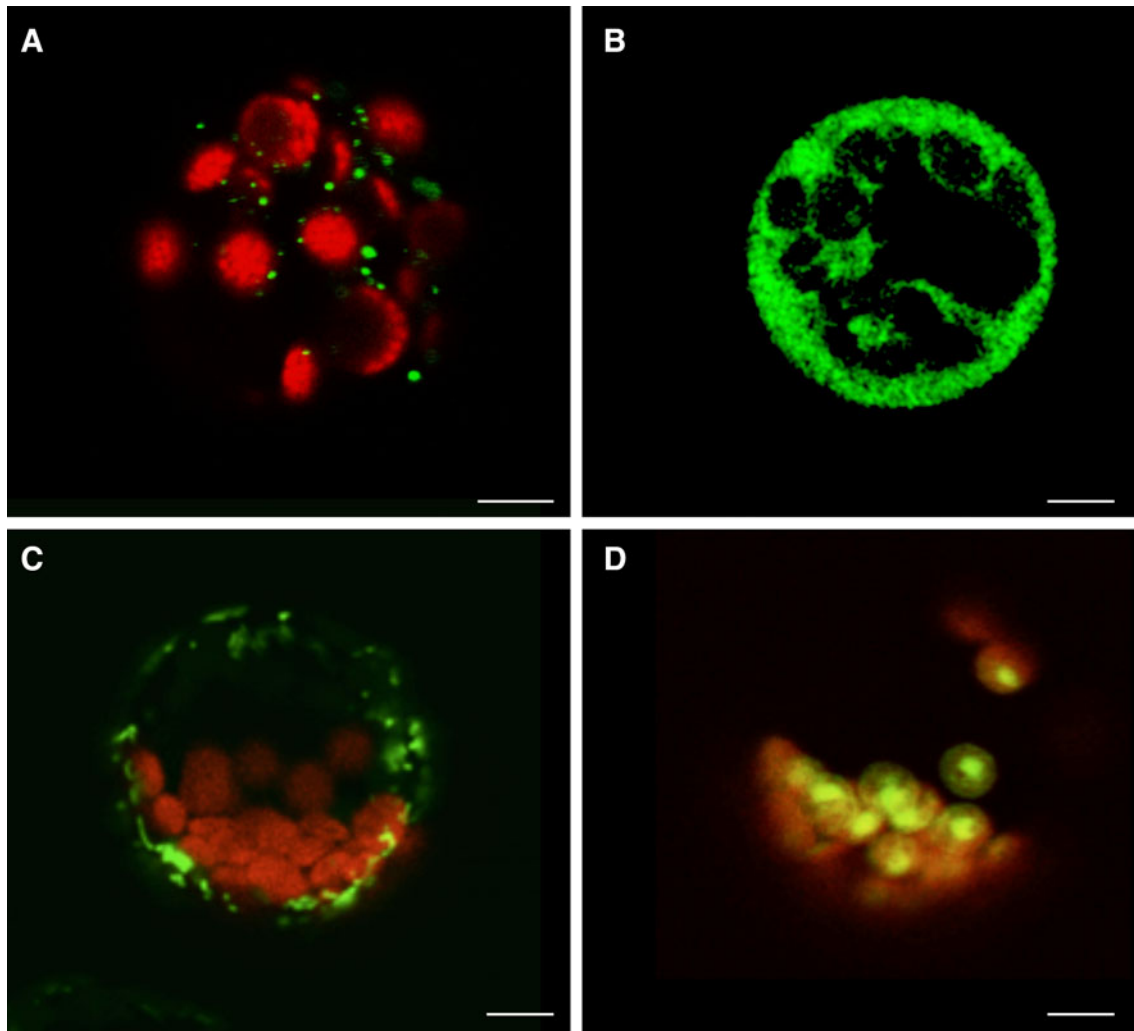


Fig. 4 Transient expression of GFP fusion proteins in *A. thaliana* protoplasts. The *PpRpoT3* gene fragment encoding a putative transit peptide was inserted into plasmid pOL-GFPS65C to generate a vector driving the expression of utrPpRpoT3 (**a**) showing mitochondrial GFP

localisation; RpoT3 was therefore designated PpRpoTm. Control constructs code for GFP (**b**, pOL only), mitochondrial CoxIV-GFP (**c**) and plastidial RecA-GFP (**d**), respectively. Images were taken by confocal fluorescence microscopy. Scale bar 10 μ m

RPOTmp2. Interestingly, the 46–48 % identity scores of RPOT3 with the three *Arabidopsis* RPOTs are higher than those of RPOTmp1 (43–47 %) and RPOTmp2 (43–46 %). Figure 1 shows an alignment of the *Arabidopsis* and *Physcomitrella* RPOT polymerases demonstrating the high degree of similarity and the conservation of functionally essential residues and motifs in PpRPOT3. Thus, the derived amino acid sequence is consistent with *PpRpoT3* encoding a functional RNA polymerase. To elucidate the gene structure of the third *Physcomitrella* *RpoT* gene, the cDNA was aligned with the genomic sequence taken from the JGI genome browser (<http://genome.jgi-psf.org>). The two published *RpoT* genes *RpoTmp1* (7.3 kb) and *RpoTmp2* (8.7 kb) contain 18 and 20 introns, respectively (Richter et al. 2002). In *PpRpoT3* (7.3 kb), only 17 introns

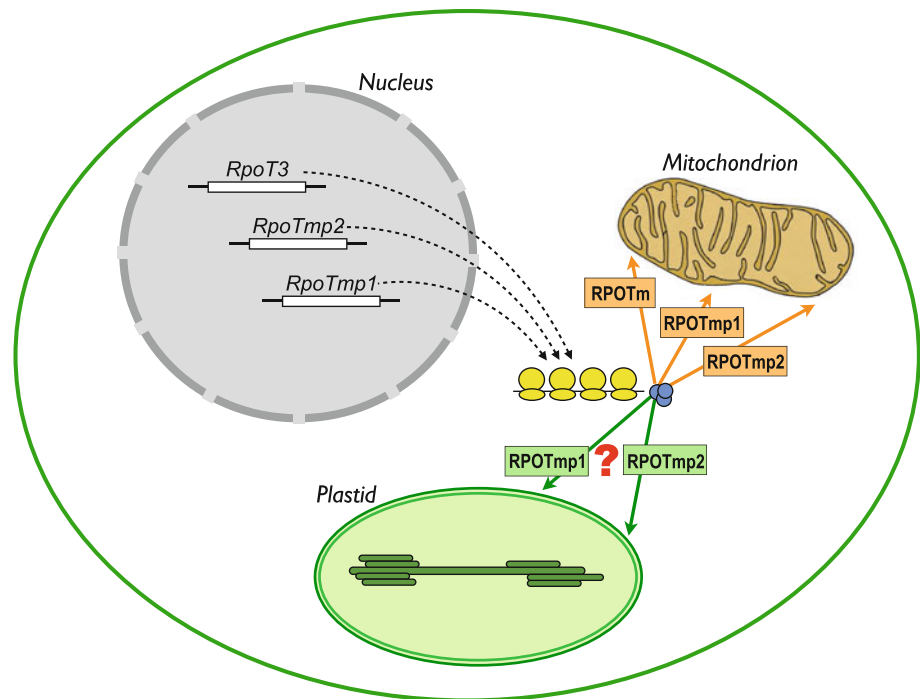
are found. Compared with the *RpoT* gene structure of flowering plants, *PpRpoT3* contains one additional intron in the 5' non-coding part of the gene, intron 4 and 16 (as of *Arabidopsis RpoTm*) are missing in *PpRpoT3* (Fig. 2).

The insertion sites of all other common introns are precisely conserved relative to the aligned amino acids sequences between *PpRpoT3*, the two other *Physcomitrella* genes and those of *Arabidopsis* (Fig. 2).

Phylogenetic analysis

No conclusive evidence concerning the phylogenetic relationship of the three *Physcomitrella RpoT* genes can be inferred from the exon–intron structure. While the acquisition of intron 1 (as of *PpRpoTmp1*) is in favour of

Fig. 5 Proposed localisation of the three phage-type RNA polymerases in *Physcomitrella* cells. Based on the results of the present study. RPOT3 is localised exclusively to mitochondria and was, therefore, renamed into RPOTm. RPOTmp1 and RPOTmp2 are potentially targeted to both mitochondria and plastids; their localisation might be regulated via usage of different start codons during translation and vary between different tissues



a closer relationship of *PpRpoTmp1* and *PpRpoTmp2*, *PpRpoTmp1* and *PpRpoT3* both share the lack of the highly conserved intron 16 (as of *Arabidopsis RpoTm*). To assess the phylogenetic positioning of the polymerase characterised in the present study more specifically, phylogenetic trees were reconstructed applying the Bayesian algorithm as well as maximum-likelihood (PhyML) analysis. Tree reconstruction was based on a multiple alignment of 54 plant RPOT sequences (see Fig. 3; Supplementary Table 2). Bayesian as well as PhyML analysis resulted in essentially the same tree topology (not shown). Figure 3 shows the rooted (green algae) consensus tree of the PhyML analysis in which angiosperm RPOT polymerases constitute two clearly discernible groups: one consisting of plastid-localised polymerases, and the other of mitochondrial-localised and dual-targeted enzymes, while *Selaginella* and *Physcomitrella* polymerases do not belong to these clusters. All *Physcomitrella* RPOTs appear to be the result of two *Physcomitrella* lineage specific duplication events, with the *PpRpoTm/RpoTmp1* duplication as the more recent one. Furthermore, conservation of the plastid-targeting properties for both dual-targeted proteins or, alternatively, independent acquisition of plastid-targeting properties for both RPOTmp1 and RPOTmp2 is supportive for their functional importance in *Physcomitrella*. The lack of a dual-targeted RPOTmp both in *Nuphar* and in monocots supports the view that the RPOTmp enzyme detected in eudicots is an evolutionary novelty due to an *RpoTm* gene

duplication that likely occurred only after the separation of monocots and eudicots. The putative plastid-targeting sequences as present in two of the three *Physcomitrella* RPOT proteins are, therefore, clearly species- or lineage-specific convergent inventions. Importantly, multiple mitochondrial RNA polymerases as found in *Physcomitrella* and eudicots are identified in *Nuphar* as well. The fixation of duplicated *RpoT* genes led to a convergent multiplicity of mitochondrial RNAPs in *Nuphar*, *Physcomitrella* and eudicots, previously not found in any other eukaryotic lineage.

Mitochondrial localisation of the phage-type RNA polymerase

While both *PpRpoTmp1* and *PpRpoTmp2* contain two in-frame AUG codons in their 5' ends, potentially facilitating dual targeting properties due to alternative translational starts, for *PpRpoT3* only one putative start codon was identified. Computer analysis of the putative transit peptide (Emanuelsson et al. 2000) produced high scores (0.8–0.9) for mitochondrial localisation and no indication for targeting to chloroplasts. To experimentally investigate the subcellular localisation of the *PpRpoT3*-encoded polypeptide, the putative N-terminal transit peptide was fused in-frame to the GFP coding region. Furthermore, an additional construct was generated in such a way that it contained the whole native 5'UTR sequence together with the sequence encoding the putative transit peptide, allowing for

translational initiation at potential non-AUG start codons and to rule out GFP localisation due to forced translational initiation (Kabeya and Sato 2005). After transfection of the *PpRpoT3*-GFP fusion constructs into *Arabidopsis* protoplasts transient expression was monitored by confocal laser scanning microscopy. Three plasmids encoding GFP only (Fig. 4b), a mitochondrial Cox4-GFP and a plastid RecA-GFP (Fig. 4d) fusion protein were used for reference transformations of *Arabidopsis* protoplasts. Protoplasts expressing PpRPOT3-GFP from both constructs displayed green fluorescence of small structures (shown in Fig. 3a for the construct with 5'UTR) resembling the fluorescent mitochondria of protoplasts synthesizing Cox4-GFP (Fig. 4c), substantiating a mitochondrial localisation of PpRPOT3. The obtained data indicates that the N-terminal part of PpRPOT3 has mitochondrial targeting properties and strongly suggests that RPOT3 is exclusively targeted to mitochondria *in planta*. Therefore, PpRPOT3 will be subsequently designated as PpRPOTm.

In the light of the existence of only one *RpoT* gene in *Selaginella*, the finding of three genes encoding organellar phage-type RNA polymerases in the nuclear genome of *P. patens* comes as a surprise (Fig. 5). It raises the question as to the functions of three RNA polymerases (PpRPOTm, PpRPOTmp1, PpRPOTmp2) in mitochondria and of possibly also three enzymes in chloroplasts (PEP and the putative NEP activities PpRPOTmp1 and PpRPOTmp2 in addition to PEP). Multiple RPOT activities per organelle appear as a common feature of eudicotyledonous plants and *Physcomitrella*. Whether this complexity is due to an evolutionary trend, adaptive or neutral, remains to be shown. The identification of different but overlapping functions of RPOT proteins in plastids as well as in mitochondria of *Arabidopsis* is only the first step to answer such a question, since a regulatory role *sensu stricto* for these enzymes remains to be shown (Liere et al. 2011). *Physcomitrella RpoT* deletion mutants and *Physcomitrella* lines with manipulated higher or lower RPOT activities might offer a way to show if a similar division of labour evolved in case of the three mitochondrial RNA polymerases in *Physcomitrella* as compared to the two *Arabidopsis* enzymes (Kühn et al. 2009).

Acknowledgments This work was supported by grants of the DFG to T.B. and A.W. (SFB 429, WE 1595/6-2). We are thankful to Chang Yin (Institute of Biology-Genetics, Humboldt University, Berlin, Germany) for her support with targeting experiments and to Ian Small (Plant Energy Biology, ARC Centre of Excellence, The University of Western Australia, Perth, Australia) for generously providing GFP fusion constructs.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Burger G, Gray MW, Forget L, Lang BF (2013) Strikingly bacteria-like and gene-rich mitochondrial genomes throughout Jakobid protists. *Genome Biol Evol* 5:418–438
- Chang CC, Sheen J, Bligny M, Niwa Y, Lerbs-Mache S, Stern DB (1999) Functional analysis of two maize cDNAs encoding T7-like RNA polymerases. *Plant Cell* 11:911–926
- Cove D (2005) The moss *Physcomitrella patens*. *Annu Rev Genet* 39:339–358
- Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, Heled J, Kearse M, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2010) Geneious v5.1, available from <http://www.geneious.com/>
- Emanuel C, Weihe A, Graner A, Hess WR, Börner T (2004) Chloroplast development affects expression of phage-type RNA polymerases in barley leaves. *Plant J* 38:460–472
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* 300:1005–1016
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321
- Hedtke B, Börner T, Weihe A (2000) One RNA polymerase serving two genomes. *EMBO Rep* 1:435–440
- Hedtke B, Legen J, Weihe A, Herrmann RG, Börner T (2002) Six active phage-type RNA polymerase genes in *Nicotiana tabacum*. *Plant J* 30:625–637
- Ikeda TM, Gray MW (1999) Identification and characterization of T3/T7 bacteriophage-like RNA polymerase sequences in wheat. *Plant Mol Biol* 40:567–578
- Kabeya Y, Sato N (2005) Unique translation initiation at the second AUG codon determines mitochondrial localization of the phage-type RNA polymerases in the moss *Physcomitrella patens*. *Plant Physiol* 138:369–382
- Kabeya Y, Hashimoto K, Sato N (2002) Identification and characterization of two phage-type RNA polymerase cDNAs in the moss *Physcomitrella patens*: implication of recent evolution of nuclear-encoded RNA polymerase of plastids in plants. *Plant Cell Physiol* 43:245–255
- Kobayashi Y, Dokiya Y, Sugita M (2001a) Dual targeting of phage-type RNA polymerase to both mitochondria and plastids is due to alternative translation initiation in single transcripts. *Biochem Biophys Res Commun* 289:1106–1113
- Kobayashi Y, Dokiya Y, Sugiura M, Niwa Y, Sugita M (2001b) Genomic organization and organ-specific expression of a nuclear gene encoding phage-type RNA polymerase in *Nicotiana sylvestris*. *Gene* 279:33–40
- Kobayashi Y, Dokiya Y, Kumazawa Y, Sugita M (2002) Non-AUG translation initiation of mRNA encoding plastid-targeted phage-type RNA polymerase in *Nicotiana sylvestris*. *Biochem Biophys Res Commun* 299:57–61
- Kühn K, Richter U, Meyer EH, Delannoy E, de Longevialle AF, O'Toole N, Börner T, Millar AH, Small ID, Whelan J (2009) Phage-type RNA polymerase RPOTmp performs gene-specific transcription in mitochondria of *Arabidopsis thaliana*. *Plant Cell* 21:2762–2779
- Liere K, Börner T (2007) Transcription and transcriptional regulation in plastids. In: Bock R (ed) *Topics in current genetics: cell and molecular biology of plastids*. Springer, Berlin/Heidelberg, pp 121–174
- Liere K, Weihe A, Börner T (2011) The transcription machineries of plant mitochondria and chloroplasts: composition, function, and regulation. *J Plant Physiol* 168:1345–1360

- McAllister WT, Raskin CA (1993) The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. *Mol Microbiol* 10:1–6
- Peeters N, Chapron A, Giritch A, Grandjean O, Lancelin D, Lhomme T, Vivrel A, Small I (2000) Duplication and quadruplication of *Arabidopsis thaliana* cysteinyl- and asparaginyl-tRNA synthetase genes of organellar origin. *J Mol Evol* 50:413–423
- Rensing SA et al (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319:64–69
- Richter U, Kiessling J, Hedtke B, Decker E, Reski R, Börner T, Weihe A (2002) Two *RpoT* genes of *Physcomitrella patens* encode phage-type RNA polymerases with dual targeting to mitochondria and plastids. *Gene* 290:95–105
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 12:1572–1574
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Sousa R, Chung YJ, Rose JP, Wang BC (1993) Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature* 364:593–599
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of multiple sequence alignment through sequence weighting, position specific gap-penalties and weight matrix choice. *Nucl Acids Res* 37:7509–7518
- Yin C, Richter U, Börner T, Weihe A (2009) Evolution of phage-type RNA polymerases in higher plants: characterization of the single phage-type RNA polymerase gene from *Selaginella moellendorffii*. *J Mol Evol* 68:528–538
- Yin C, Richter U, Börner T, Weihe A (2010) Evolution of plant phage-type RNA polymerases: the genome of the basal angiosperm *Nuphar advena* encodes two mitochondrial and one plastid phage-type RNA polymerases. *BMC Evol Biol* 10:379
- Yoo SD, Cho YH, Sheen J (2007) *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat Protoc* 2:1565–1572