ORIGINAL PAPER

Analyzing drivers of speciation in the Southern Ocean using the sea spider species complex *Colossendeis megalonyx* as a test case

Jana S. Dömel[1] · Lars Dietz[2,3] · Till-Hendrik Macher[1] · Andrey Rozenberg[4] · Christoph Mayer[2,3] · Johanna M. Spaak[5] · Roland R. Melzer[6,7,8] · Florian Leese[1,9]

Abstract

Colossendeis megalonyx Hoek, 1881 has the broadest distribution of all sea spiders in the Southern Ocean. Previous studies have detected several evolutionarily young lineages within this taxon and interpreted them as a result of allopatric speciation in a few shelf refuges during glacial maxima. However, alternative scenarios such as ecological speciation in sympatry have rarely been considered or tested. Here, we generated the most extensive genomic and morphometric data set on the *C. megalonyx* species complex to (i) comprehensively describe species diversity, (ii) explore intraspecific connectivity between populations located around Antarctica, and (iii) systematically test for positive selection indicative of adaptive speciation. We successfully applied a target hybrid enrichment approach and recovered all 1607 genes targeted. Phylogenomic analysis was consistent with previous findings and, moreover, increased the resolution of branching within lineages. We found specimens of phylogenetically well-separated lineages occurring in sympatry to be genetically distinct from each other and gene flow between geographically separated populations of the same lineages to be restricted. Evidence for positive selection was found for four genes associated with structural and neuronal functions. Hence, there is an indication for positive selection in the *C. megalonyx* species complex, yet its specific contribution to the speciation process remains to be explored further. Finally, morphometric analyses revealed multiple significant differences between lineages, but a clear separation proved difficult. Our study highlights the relevance of positive selection as a potential driver for speciation in the Southern Ocean.

Keywords Antarctic benthos · Target hybrid enrichment · Positive selection · Integrative taxonomy · Cryptic species

Introduction

The Southern Ocean accommodates a unique, speciose and highly endemic benthic community (Knox and Lowry 1977; Clarke and Johnston 2003; Aronson et al. 2007; Clarke 2008) that has evolved through in situ radiations after the

✉ Jana S. Dömel
jana.doemel@uni-due.de

1 Aquatic Ecosystem Research, Faculty of Biology, University of Duisburg-Essen, Universitaetsstr. 5, 45141 Essen, Germany

2 Statistical Phylogenetics and Phylogenomics, Zoological Research Museum Alexander Koenig, Adenauerallee 160, 53113 Bonn, Germany

3 Faculty of Mathematics and Natural Sciences, University of Bonn, 53012 Bonn, Germany

4 Faculty of Biology, Technion –Israel Institute of Technology, 3200003 Haifa, Israel

5 Biological Sciences, School of Science, University of Waikato, Hamilton 3216, New Zealand

6 Bavarian State Collection of Zoology - SNSB, Muenchhausenstr. 21, 81247 Munich, Germany

7 Department Biologie II, Ludwig Maximilians University of Munich, Großhaderner Str. 2, 82152 Planegg-Martinsried, Germany

8 GeoBioCenter, LMU Munich, Richard-Wagner-Str. 10, 80333 Munich, Germany

9 Centre for Water and Environmental Research (ZWU), University of Duisburg-Essen, Universitaetsstr. 2, 45141 Essen, Germany

opening of the Drake Passage (about 30 mya) and the subsequent establishment of the Polar Front (see Poulin et al. 2002; Briggs 2003; Convey et al. 2009). Prominent examples of early in situ radiations during the Mio- and Pliocene include the Antarctic icefishes (Notothenioidei, Eastman and McCune 2000), octopuses (Strugnell et al. 2012), and serolid isopods (Brandt 1991; Held 2000). In addition to these ancient radiations, molecular studies revealed a high number of recent divergences that probably occurred in the Plio- and Pleistocene (e.g., Convey et al. 2009; Fraser et al. 2012; Halanych and Mahon 2018 for overviews). For these, recurrent glaciations are seen as important drivers of speciation (Clarke and Crame 1989; Thatje et al. 2005). Grounded shelf ice covered major areas of the continental shelf during glacial maxima, thereby making it uninhabitable for benthic communities (Convey et al. 2009). At the same time, recent evidence suggests that small ice-free refugia existed and that relict populations could survive on the continental shelf (see Thatje et al. 2005, 2008 and Fraser et al. 2012, 2014 for overviews). The assumed limited dispersal potential of many Southern Ocean benthic species in combination with the higher rates of random genetic drift in small, isolated populations, likely triggered rapid lineage sorting in independently evolving populations and ultimately allopatric speciation. The often allopatric occurrence of cryptic species adds strong support for this hypothesis (e.g., Held 2003; Held and Wägele 2005; Wilson et al. 2007).

However, several studies have documented that divergent selection also has a large contribution to speciation processes (Schluter 2000; Dieckmann et al. 2004; Coyne and Orr 2004), for example, in cichlid fishes in African and central American lakes (e.g., Kocher 2004), or the marine snail *Littorina saxatilis* Olivi, 1792 (Johannesson et al. 2017). Interestingly, the role of selection in Southern Ocean speciation processes has rarely been addressed. Rutschmann et al. (2011) tested for evidence of ecological specialization underlying the adaptive radiation of notothenioid fishes and found a lineage-independent, ecological differentiation into different niches. A study on the Southern Ocean sea slug *Doris kerguelenensis* (Bergh, 1884) reported evidence that interspecific competition (predation) was involved in more recent speciation during the Plio- and Pleistocene (Wilson et al. 2009). More intense analyses including more markers found distinct anti-predatory secondary metabolites in different cryptic lineages, thereby supporting the hypothesis that defense against predation was a (differential) mechanism leading to adaptive speciation (Wilson et al. 2013). Those results already show that recently diverged species of the Southern Ocean benthos represent suitable evolutionary test cases to address the possible impacts of selection in lineage diversification.

One taxon consisting of a remarkable amount of recently diverged lineages are pycnogonids (Mahon et al.

2008; Krabbe et al. 2010; Arango et al. 2011; Weis and Melzer 2012; Weis et al. 2014; Dietz et al. 2015, 2019; Dömel et al. 2015, 2017; Ballesteros et al. 2020). Pycnogonids, also called sea spiders, are a group of exclusively marine arthropods that are especially diverse in the Southern Ocean (Aronson et al. 2007), showing high endemism to this region (Munilla and Soler-Membrives 2009). Thus, sea spiders are characteristic representatives of the Southern Ocean benthos suitable to study drivers of recent speciation processes under the extreme conditions in Antarctica. A prominent example of a sea spider species complex is *Colossendeis megalonyx* Hoek, 1881. *Colossendeis megalonyx* is one of the most broadly distributed sea spiders in the Southern Ocean (Griffiths et al. 2011; Dietz et al. 2015). It occurs in shallow and deep Antarctic and sub-Antarctic benthic habitats as well as around South America, South Africa, and Madagascar (Munilla and Soler-Membrives 2009). The feeding spectrum seems to be broad and even pelagic invertebrates are frequently consumed (Moran et al. 2018). However, due to the fact that many cryptic species are known in that group, investigations of food preference have to be specified for each lineage before the abovementioned observations can be generalized (Dietz et al. 2018). Furthermore, the mode of reproduction remains a mystery, as no larval stages have ever been reported for *Colossendeis* and even for all Colossendeidae (Arnaud and Bamber 1987). Previous studies found several cryptic lineages within *C. megalonyx* (Krabbe et al. 2010; Dietz et al. 2015). Depending on the genetic markers those studies looked at, the proposed number of delineated lineages ranged from 15 to 20 (clades A to O) for the mitochondrial cytochrome c oxidase subunit I gene (COI) and six (I–VI) for the highly variable nuclear internal transcribed spacer region (ITS) for over 300 analyzed specimens (Dietz et al. 2015). Also, the nuclear gene for histone H3 was tested for a subsample of specimens and revealed similar groupings as proposed for ITS (Krabbe 2010). At the same time, comparisons of trees reconstructed from nuclear versus mitochondrial markers revealed mito-nuclear discordance in *C. megalonyx* (Dietz et al. 2015). This was explained by speciation reversals induced by hybridization between members of the formerly distinct mitochondrial clades. While recombination leads to gradual homogenization of nuclear genomes in hybridizing lineages, the deep mitochondrial divergences remain due to the absence of recombination in mitochondrial DNA. Distribution ranges differed between clades with some being local, but many others having a circumpolar distribution, and members of several lineages occur in sympatry (Dietz et al. 2015). Here the questions arise, how is co-existence maintained, and can ecological character displacement to minimize interspecific competition and allow for stable co-existence be expected. Previous

work already reported morphological differences for some of the mitochondrial clades, e.g., specimens of clade C lack pigmented eyes, and specimens of clade F are larger than others (Krabbe et al. 2010; Dietz et al. 2015). Hence, *C. megalonyx* offers an ideal test case to investigate the impact of ecological divergence and the action of positive, divergent selection for recently diverged species that occur in sympatry.

New genomic tools allow to test for evidence of neutral versus adaptive divergence. With the establishment of high-throughput sequencing technologies and bioinformatics programs, these tools become accessible even for non-model species (e.g., Faircloth et al. 2012; Lemmon et al. 2012; Hugall et al. 2015; Mayer et al. 2016; Weiss et al. 2018; Breinholt et al. 2018). Although full genome sequencing followed by annotation and detailed analysis is the ideal method to perform genomic analyses and test for selection (e.g., Ellegren 2014; Nater et al. 2015), the high costs for high-quality genomes of multiple samples limit its application. Other, less expensive approaches are transcriptomic analyses (e.g., Morin et al. 2008; Lemer et al. 2015), yet the downside for studies on organisms from remote marine habitats is that often no material with well-preserved RNA is available. A suitable alternative is target hybrid enrichment (Faircloth et al. 2012), which allows capturing hundreds of target genes by hybridizing genomic DNA against specifically designed probes, the so-called baits. Another advantage of target hybrid enrichment is that low quality and quantity of DNA are sufficient, and therefore, the method can also be applied to samples with degenerated DNA (Mayer et al. 2016; Wood et al. 2018). This makes the technique a good candidate to be applied for Southern Ocean benthos studies where the limitation of well-preserved material is obvious. First studies have shown that the technique can offer new insights into the phylogeny and diversification of non-target species (e.g., Abdelkrim et al. 2018; O'Hara et al. 2019). Also population genetic patterns within lineages can be assessed especially when including co-enriched flanking regions of targeted genes (Dömel et al. 2019). Furthermore, the method explicitly targets coding regions, and sequences of these can be analyzed for signatures of selection, e.g., through dN/dS tests (Teasdale et al. 2016).

In this study, we examined the power of target hybrid enrichment to explore drivers of speciation in the recently diverged *C. megalonyx* sea spider species complex. Specifically, we aimed to (i) comprehensively describe species diversity within the complex, which is still under debate, (ii) explore genetic connectivity among populations for a subset of lineages, and (iii) test for positive selection in the different lineages. Data were complemented by new morphological data to search for significant differences between molecularly separated lineages and to explore whether morphological characters also hint at selection-favoring divergent phenotypes of genetically distinct species in sympatry. Finally, results were compared to a study of another sea spider species complex, *Pallenopsis patagonica* (Hoek, 1881), that addressed similar questions (Dömel et al. 2019).

Materials and methods

Materials

For genetic analyses of *Colossendeis megalonyx,* a subset of specimens analyzed by Dietz et al. (2015) was selected, including individuals from the Antarctic continental shelf and waters around sub-Antarctic and Patagonian islands (Table 1). Furthermore, new samples from the South Sandwich Islands were included. Specimens belonged to seven different mitochondrial clades (A–F and I) and five nuclear groups (I–IV and VI) after Dietz et al. (2015). To test if sympatrically occurring lineages show marked differences in genes due to positive selection, we focused especially on specimens from the two mitochondrial clades D1 and E1, which are reported to be sympatric in East Antarctica (Dietz et al. 2015). Hence, 50 individuals of those two clades from three different locations were included in this study. In total, 64 samples were analyzed using target hybrid enrichment (Fig. 1).

We also included morphological measurements of 103 specimens from ten mitochondrial clades. Here, in addition to the genetic data set, specimens of the clades G, N, and O were analyzed. Twenty of these specimens were also included in the genetic analysis.

Methods

Target hybrid enrichment

For target hybrid enrichment, a specific bait set was designed for *C. megalonyx* following the workflow described by Mayer et al. (2016). The bait design was based on genes that are single copy in spider genomes and that are present in a transcriptome assembly of *C. megalonyx* (see Dietz et al. 2019 for details and bait sequences). The bait set included a total number of 12,014 baits covering 3682 bait regions from 1607 eukaryotic orthologous genes (EOGs). Baits and target enrichment kit were ordered from Agilent Technologies (Waldbronn, Germany).

For sample preparation, DNA extracts from Dietz et al. (2015) were used when possible (see methods in Dietz et al. 2015 for DNA extraction details). For DNA samples with low quantity (for enrichment 100 ng per 10 µl sample were needed) and newly included individuals, DNA was extracted following the salt precipitation protocol after Sunnucks and Hales (1996; modified as in Weiss and Leese 2016). For

Table 1 Specimens of the *Colossendeis megalonyx* species complex used during study

ID	Location	Lat	Lon	Depth [m]	Cruise	COI clade	ZSM-Voucher number	COI GenBank number	Genomic analyses	Morphometric analyses
#87-21	South Georgia	−54.22	−36.53	230	ICEFISH	A	ZSM-A20171215	KT201823		x
41BT22-1	Bouvet Island	−54.64	3.30	648	ICEFISH	A	ZSM-A20171302	KT201815	x	x
76OT50-1	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171231	KT202070		x
76OT50-2	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171230	KT202068		x
ACE2017_830_1_3	South Sandwich Islands	−57.16	−26.79	326	ACE 2017	A			x	
ACE2017_962_1_1_2	South Sandwich Islands	−59.47	−27.33	130	ACE 2017	E1	ZSM-A20171329	KT202010	x	x
AGT42/164-1	South Shetland Islands	−62.13	−57.67	555	PS42	D1	ZSM-A20171352	KT202131		x
AGT42/164-4	South Shetland Islands	−62.13	−57.67	555	PS42	N3	ZSM-A20171298			x
AGT42/175-1	South Shetland Islands	−62.32	−58.70	496	PS42	F	ZSM-A20171260	KT202139	x	x
AGT42/175-2	South Shetland Islands	−62.32	−58.70	496	PS42	O	ZSM-A20171258	KT201964		x
AGT42/175-5	South Shetland Islands	−62.32	−58.70	496	PS42	F	ZSM-A20171259	KT202140		x
AGT42/175-8	South Shetland Islands	−62.32	−58.70	496	PS42	O	ZSM-A20171309			x
AGT42/175-9	South Shetland Islands	−62.32	−58.70	496	PS42	D1		KT202033		x
MNHN-IU-2007-130	Terre Adelie	−67.00	145.30	1352	CEAMARC	E1		KT202089	x	x
MNHN-IU-2007-248	Terre Adelie	−65.70	140.57	545	CEAMARC	E1		KT202060	x	
MNHN-IU-2007-249	Terre Adelie	−65.70	140.57	545	CEAMARC	D1		KT202090	x	
MNHN-IU-2014-20292	Terre Adelie	−65.70	140.57	545	CEAMARC	E1		KT202046	x	
MNHN-IU-2007-255	Terre Adelie	−66.34	139.99	568	CEAMARC	D1		KT202100		
MNHN-IU-2007-287	Terre Adelie	−66.75	145.33	604	CEAMARC	E1		KT202047		x
MNHN-IU-2014-20293	Terre Adelie	−65.87	144.11	1096	CEAMARC	D1		KT202096	x	
MNHN-IU-2014-20294	Terre Adelie	−65.87	144.11	1096	CEAMARC	E1		KT202097	x	
MNHN-IU-2014-20295	Terre Adelie	−65.87	144.11	1096	CEAMARC	E1		KT202094	x	
MNHN-IU-2014-20296	Terre Adelie	−65.87	144.11	1096	CEAMARC	E1		KT202048	x	
MNHN-IU-2014-20297	Terre Adelie	−65.87	144.11	1096	CEAMARC	D1		KT202092	x	
MNHN-IU-2014-20298	Terre Adelie	−65.87	144.11	1096	CEAMARC	E1		KT202104	x	
MNHN-IU-2014-20299	Terre Adelie	−65.87	144.11	1096	CEAMARC	E1		KT202095	x	
MNHN-IU-2014-20301	Terre Adelie	−65.47	139.36	799	CEAMARC	E1		KT202098	x	
MNHN-IU-2007-39	Terre Adelie	−66.00	143.00	470	CEAMARC	E1		KT201965	x	
MNHN-IU-2007-41	Terre Adelie	−66.45	140.53	1204	CEAMARC	F		KT202049		x
MNHN-IU-2007-50	Terre Adelie	−66.00	142.66	447	CEAMARC	D1		KT202105	x	
MNHN-IU-2007-51	Terre Adelie	−66.00	142.66	447	CEAMARC	E1		KT202108	x	
MNHN-IU-2007-62	Terre Adelie	−66.33	143.27	698	CEAMARC	E1		KT202091	x	
MNHN-IU-2014-20302	Terre Adelie	−66.33	143.27	698	CEAMARC	E1		KT202101	x	
MNHN-IU-2014-20303	Terre Adelie	−66.33	143.27	698	CEAMARC	E1		KT202093	x	
MNHN-IU-2007-73	Terre Adelie	−66.75	145.00	648	CEAMARC	E1			x	
PA_E005	South Sandwich Islands	−57.06	−26.75	130	ICEFISH	A		GQ387009		x

Table 1 (continued)

ID	Location	Lat	Lon	Depth [m]	Cruise	COI clade	ZSM-Voucher number	COI GenBank number	Genomic analyses	Morphometric analyses
PA_E010	South Sandwich Islands	−58.20	−26.10	400	ICEFISH	E1	ZSM-A20171305	GQ387026		x
PB_E002	South Sandwich Islands	−58.20	−26.10	400	ICEFISH	E1	ZSM-A20171266	GQ387026	x	x
PF_E001	Bouvet Island	−54.64	3.30	78	ICEFISH	E1	ZSM-A20171267	GQ387026		x
PF_E002	Bouvet Island	−54.64	3.30	648	ICEFISH	C	ZSM-A20171286	GQ387017	x	x
PF_E003	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171288	GQ387026	x	x
PF_E004	Bouvet Island	−54.64	3.30	648	ICEFISH	E1		GQ387026		x
PF_E005	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171287	GQ387026	x	x
PF_E006	Bouvet Island	−54.64	3.30	648	ICEFISH	C	ZSM-A20171210	GQ387016		
PF_E007	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171335	GQ387026	x	x
PF_E009	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171334	GQ387026	x	
PF_E010	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171333	GQ387026	x	
PF_E011	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171350	GQ387026		x
PG_E003	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171269	GQ387026	x	x
PG_E005	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171270	GQ387026	x	
PG_E006	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171271	GQ387026	x	
PG_E009	Bouvet Island	−54.64	3.30	648	ICEFISH	C	ZSM-A20171211	GQ387016		x
PG_E010	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171285	GQ387026	x	x
PG_E011	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171336	GQ387026	x	
PH_E004	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171289	GQ387026	x	
PH_E005	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171342	GQ387026		x
PH_E007	Bouvet Island	−54.64	3.30	648	ICEFISH	C	ZSM-A20171277	GQ387017	x	x
PH_E011	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171339	GQ387026		x
PH_E012	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171290	GQ387026		x
PI_E002	Bouvet Island	−54.64	3.30	648	ICEFISH	C		GQ387016		x
PI_E003	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171351	GQ387026		x
PI_E005	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171301	GQ387026	x	x
PJ_E001	South Shetland Islands	−61.20	−56.02	148	ICEFISH	D1		GQ387025		x
PJ_E002	South Shetland Islands	−61.20	−56.02	148	ICEFISH	D1	ZSM-A20171321	GQ387024		x
PJ_E003	South Shetland Islands	−61.20	−56.02	148	ICEFISH	D1	ZSM-A20171313	GQ387025		x
PJ_E004	South Shetland Islands	−61.20	−56.02	148	ICEFISH	D1	ZSM-A20171322	GQ387025		x
PJ_E005	South Shetland Islands	−61.20	−56.02	148	ICEFISH	D1	ZSM-A20171323	GQ387025		x
PJ_E006	Burdwood Bank	−54.78	−59.25	300	ICEFISH	B	ZSM-A20171311	GQ387013		x
PJ_E007	Burdwood Bank	−54.78	−59.25	300	ICEFISH	B	ZSM-A20171310	GQ387013		x
PJ_E009	Burdwood Bank	−54.78	−59.25	300	ICEFISH	B	ZSM-A20171312	GQ387013		x

Table 1 (continued)

ID	Location	Lat	Lon	Depth [m]	Cruise	COI clade	ZSM-Voucher number	COI GenBank number	Genomic analyses	Morphometric analyses
PL_E004	South Sandwich Islands	−58.95	−26.45	120	ICEFISH	A	ZSM-A20171220	GQ387007		x
PM_E008	South Shetland Islands	−62.32	−58.70	491	ICEFISH	F	ZSM-A20171256	GQ387027		x
PM_E010	South Shetland Islands	−62.32	−58.70	496	ICEFISH	D1	ZSM-A20171326	GQ387019		x
PM_E011	South Shetland Islands	−62.32	−58.70	496	ICEFISH	D1	ZSM-A20171274	GQ387023		x
PN_E002	South Shetland Islands	−61.78	−57.50	322	ICEFISH	A	ZSM-A20171216	GQ387008		x
PN_E007	South Shetland Islands	−61.77	−57.53	343	ICEFISH	A	ZSM-A20171300	GQ387007	x	x
PN_E008	South Shetland Islands	−62.32	−58.70	491	ICEFISH	D1	ZSM-A20171324	GQ387019		x
PN_E009	South Shetland Islands	−62.32	−58.70	496	ICEFISH	D1	ZSM-A20171275	GQ387022		x
PN_E010	South Shetland Islands	−62.32	−58.70	491	ICEFISH	D1	ZSM-A20171282	GQ387020	x	x
PN_E011	South Shetland Islands	−62.32	−58.70	491	ICEFISH	D1	ZSM-A20171325	GQ387019		x
PN_E012	South Shetland Islands	−62.32	−58.70	491	ICEFISH	F	ZSM-A20171222			x
PO_E003	Burdwood Bank	−52.67	−60.27	202	ICEFISH	B	ZSM-A20171319	GQ387013		x
PO_E004	Burdwood Bank	−52.67	−60.27	202	ICEFISH	B		GQ387015		x
PO_E005	Burdwood Bank	−52.67	−60.27	202	ICEFISH	B	ZSM-A20171296	GQ387011	x	x
PO_E006	Burdwood Bank	−52.66	−60.29	202	ICEFISH	B	ZSM-A20171320			x
PO_E007	Burdwood Bank	−52.67	−60.27	202	ICEFISH	B	ZSM-A20171317	GQ387010		x
PO_E008	Burdwood Bank	−52.67	−60.27	202	ICEFISH	B	ZSM-A20171318	GQ387014		x
PO_E009	Burdwood Bank	−52.67	−60.27	202	ICEFISH	B	ZSM-A20171315	GQ387014		x
PO_E010	Burdwood Bank	−52.67	−60.27	202	ICEFISH	B	ZSM-A20171253	GQ387013		x
PO_E011	Burdwood Bank	−52.67	−60.27	202	ICEFISH	B	ZSM-A20171316	GQ387010		x
PO_E012	Burdwood Bank	−52.67	−60.27	202	ICEFISH	B	ZSM-A20171297	GQ387012		x
PP_E001	Burdwood Bank	−52.67	−60.27	202	ICEFISH	B	ZSM-A20171314	GQ387013		x
PP_E002	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171280	GQ387026	x	
PP_E003	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171338	GQ387026	x	x
PP_E004	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171284	GQ387026		x
PP_E005	Bouvet Island	−54.64	3.30	648	ICEFISH	C	ZSM-A20171207	GQ387016		x
PP_E006	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171283	GQ387026	x	x
PP_E009	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171281	GQ387026		x
PQ_E001	Bouvet Island	−54.64	3.30	648	ICEFISH	C	ZSM-A20171276	GQ387017		x
PQ_E002	Bouvet Island	−54.64	3.30	648	ICEFISH	C	ZSM-A20171208	GQ387016		x
PQ_E003	Bouvet Island	−54.64	3.30	648	ICEFISH	C	ZSM-A20171209	GQ387016		x
PQ_E004	Bouvet Island	−54.64	3.30	648	ICEFISH	C	ZSM-A20171206	GQ387017		x
PQ_E005	Bouvet Island	−54.64	3.30	648	ICEFISH	E1	ZSM-A20171225	GQ387026		x
PQ_E006	Bouvet Island	−54.35	3.17	458	ICEFISH	E1	ZSM-A20171226	GQ387026		x

Table 1 (continued)

ID	Location	Lat	Lon	Depth [m]	Cruise	COI clade	ZSM-Voucher number	COI GenBank number	Genomic analyses	Morphometric analyses
PQ_E009	Bouvet Island	−54.35	3.17	458	ICEFISH	E1	ZSM-A20171227	GQ387026		x
PR_E001	Bouvet Island	−54.35	3.17	458	ICEFISH	E1	ZSM-A20171229	GQ387026		x
PR_E002	Bouvet Island	−54.35	3.17	458	ICEFISH	E1	ZSM-A20171228	GQ387026		x
PS_E001	Burdwood Bank	−52.65	−59.22	129	ICEFISH	B	ZSM-A20171272	GQ387013	x	x
PS_E002	Burdwood Bank	−52.65	−59.22	129	ICEFISH	B	ZSM-A20171273	GQ387014		x
PS_E003	South Sandwich Islands	−58.95	−26.45	75	ICEFISH	A	ZSM-A20171268	GQ387007	x	x
PS_E004	South Sandwich Islands	−58.95	−26.45	75	ICEFISH	A	ZSM-A20171217	GQ387007		x
PS_E005	South Sandwich Islands	−58.95	−26.45	75	ICEFISH	A	ZSM-A20171306	GQ387007		x
PS_E006	South Sandwich Islands	−58.95	−26.45	75	ICEFISH	A		GQ387007		x
PS_E007	South Shetland Islands	−62.32	−58.70	491	ICEFISH	D1	ZSM-A20171278	GQ387018		x
PS_E008	South Shetland Islands	−62.32	−58.70	491	ICEFISH	D1	ZSM-A20171279	GQ387021		x
PS_E009	South Shetland Islands	−62.32	−58.70	491	ICEFISH	D1	ZSM-A20171332	GQ387019		x
PS61_45_3	South Shetland Islands	−60.99	−55.19	196	PS 61	A	ZSM-A20171213	KT201911		x
PS61_45_4	South Shetland Islands	−60.99	−55.19	196	PS 61	A	ZSM-A20171214	KT201912		x
PS77_208_5_2	Burdwood Bank	−54.55	−56.17	294	PS77	B	ZSM-A20171293	KT202115	x	x
PS77_257_2_5_1	Eastern Antarctic Peninsula	−64.91	−60.65	158	PS77	N3	ZSM-A20171308	KT202129		x
PS77_257_2_5_2	Eastern Antarctic Peninsula	−64.91	−60.65	158	PS77	E1	ZSM-A20171292	KT202102	x	x
PS77_260_6_14	Eastern Weddell Sea	−70.84	−10.60	252	PS77	I	ZSM-A20171294	KT201735	x	x
PS77_260_6_15	Eastern Weddell Sea	−70.84	−10.60	252	PS77	I	ZSM-A20171307	KT201749		x
PS77_260_6_17	Eastern Weddell Sea	−70.84	−10.60	252	PS77	I	ZSM-A20171295	KT201736		x
PS77_260_6_22	Eastern Weddell Sea	−70.84	−10.60	252	PS77	I	ZSM-A20171247	KT201740		x
PS77_260_6_3	Eastern Weddell Sea	−70.84	−10.60	252	PS77	G	ZSM-A20171252	KT201805		x
PS77_260_6_7	Eastern Weddell Sea	−70.84	−10.60	252	PS77	G	ZSM-A20171223	KT201807		x
PS77_260_6_8	Eastern Weddell Sea	−70.84	−10.60	252	PS77	I	ZSM-A20171240	KT201746		x
PS77_260_6_9	Eastern Weddell Sea	−70.84	−10.60	252	PS77	I	ZSM-A20171248	KT201784		x
PS77_286_1_3_2	Eastern Weddell Sea	−70.84	−10.60	248	PS77	N2	ZSM-A20171255	KT202124		x
PS77_286_1_5	Eastern Weddell Sea	−70.84	−10.60	248	PS77	F	ZSM-A20171254	KT201957		x
YPM48435-1	South Orkney Islands	−60.40	−46.76	280	Yale	I		KT201775	x	
YPM48435-2	South Orkney Islands	−60.40	−46.76	280	Yale	D1		KT201973	x	
YPM48435-3	South Orkney Islands	−60.40	−46.76	280	Yale	D1		KT202024	x	
YPM48435-4	South Orkney Islands	−60.40	−46.76	280	Yale	D1		KT201977	x	
YPM48435-5	South Orkney Islands	−60.40	−46.76	280	Yale	D1		KT202012	x	
YPM48435-7	South Orkney Islands	−60.40	−46.76	280	Yale	D1		KT201979	x	
YPM48435-8	South Orkney Islands	−60.40	−46.76	280	Yale	D1		KT201983	x	

Table 1 (continued)

ID	Location	Lat	Lon	Depth [m]	Cruise	COI clade	ZSM-Voucher number	COI GenBank number	Genomic analyses	Morphometric analyses
YPM48435-9	South Orkney Islands	−60.40	−46.76	280	Yale	D1		KT202050	x	
YPM48435-15	South Orkney Islands	−60.40	−46.76	280	Yale	D1		KT202011	x	
YPM48435-17	South Orkney Islands	−60.40	−46.76	280	Yale	D1		KT201980	x	
YPM48435-23	South Orkney Islands	−60.40	−46.76	280	Yale	D1		KT201975	x	
YPM48435-27	South Orkney Islands	−60.40	−46.76	280	Yale	D1		KT201984	x	
YPM48435-38	South Orkney Islands	−60.40	−46.76	280	Yale	D1		KT201967	x	
YPM48435-39	South Orkney Islands	−60.40	−46.76	280	Yale	D1		KT201991	x	
YPM48435-40	South Orkney Islands	−60.40	−46.76	280	Yale	D1		KT201992	x	

Sampling details (location, latitude, longitude, depth, cruise), mitochondrial clade assignment, voucher number for Bavarian State Collection (ZSM) and GenBank as well as use for genomic and/or morphometric analyses

all DNA samples, RNase digestion was conducted using 1 μl RNase A per 50 μl DNA sample. Purification was performed using the NucleoSpin Gel and PCR Cleanup kit (MACHEREY–NAGEL GmbH & Co. KG, Düren, Germany). DNA was subsequently eluted in 10 μl HPLC-grade water. 1 μl DNA was used to measure concentration with the Qubit Fluorometer using the dsDNA BR Array Kit (Thermo Fisher Scientific, Waltham, USA). For qualitative analyses, 2 μl DNA was used on a Fragment Analyzer (Agilent Technologies, Santa Clara, USA) using the Standard or High-Sensitivity Genomic DNA kits (DNF-487-33/DNF-487-33). Sample enrichment and sequencing details were performed as described in Dömel et al. (2019). Two libraries containing 32 samples each were sent to GATC Biotech GmbH (Constance, Germany) for sequencing on an Illumina MiSeq platform using the V2 2×250 bp paired-end sequencing kit. 5% PhiX spike-in was added to each run to increase sequencing diversity and hence improve the signal of sequences. Upon delivery, the NGS reads were adapter- and quality-trimmed with fastq-mcf (Aronesty 2011). Raw data are available from the NCBI Sequence Read Archive (SRA: BioProject ID PRJNA545212). In the following, we used two complementary approaches to construct data sets for different purposes (see Fig. 2 for an overview). First, EOGs only were used to infer a robust species phylogeny based on orthologous regions as well as to search for signatures of selection (dN/dS tests). Secondly, two data sets were designed using a single nucleotide polymorphisms (SNPs) approach.

Tree reconstruction and test for selection

For details about phylogenetic tree reconstruction and subsequent tests for selection, see Dömel et al. (2019). The raw reads were mapped against the sequences of the bait regions within the *C. megalonyx* transcriptome using the BWA-MEM algorithm of bwa v. 0.7.17 (available from https://bio-bwa.sourceforge.net). Reads for which the initial mapping was successful were mapped against the corresponding full coding region. From this we obtained consensus sequences for most genes and species. The consensus sequences produced in this way were already aligned to the reference sequences and no further alignments were necessary.

A maximum likelihood phylogenetic analysis was performed on the concatenated data set with IQ-TREE v. 1.5.4 (Nguyen et al. 2015) using ultrafast bootstrapping with 1000 replicates for estimating nodal support. The most likely model of evolution was selected with ModelFinder (Kalyaanamoorthy et al. 2017), a tool integrated into IQ-TREE. The alignment was partitioned by codon positions, and the optimal partitioning scheme was selected with the algorithm implemented in ModelFinder (Chernomor et al. 2016). For rooting the phylogenetic tree, enrichment data
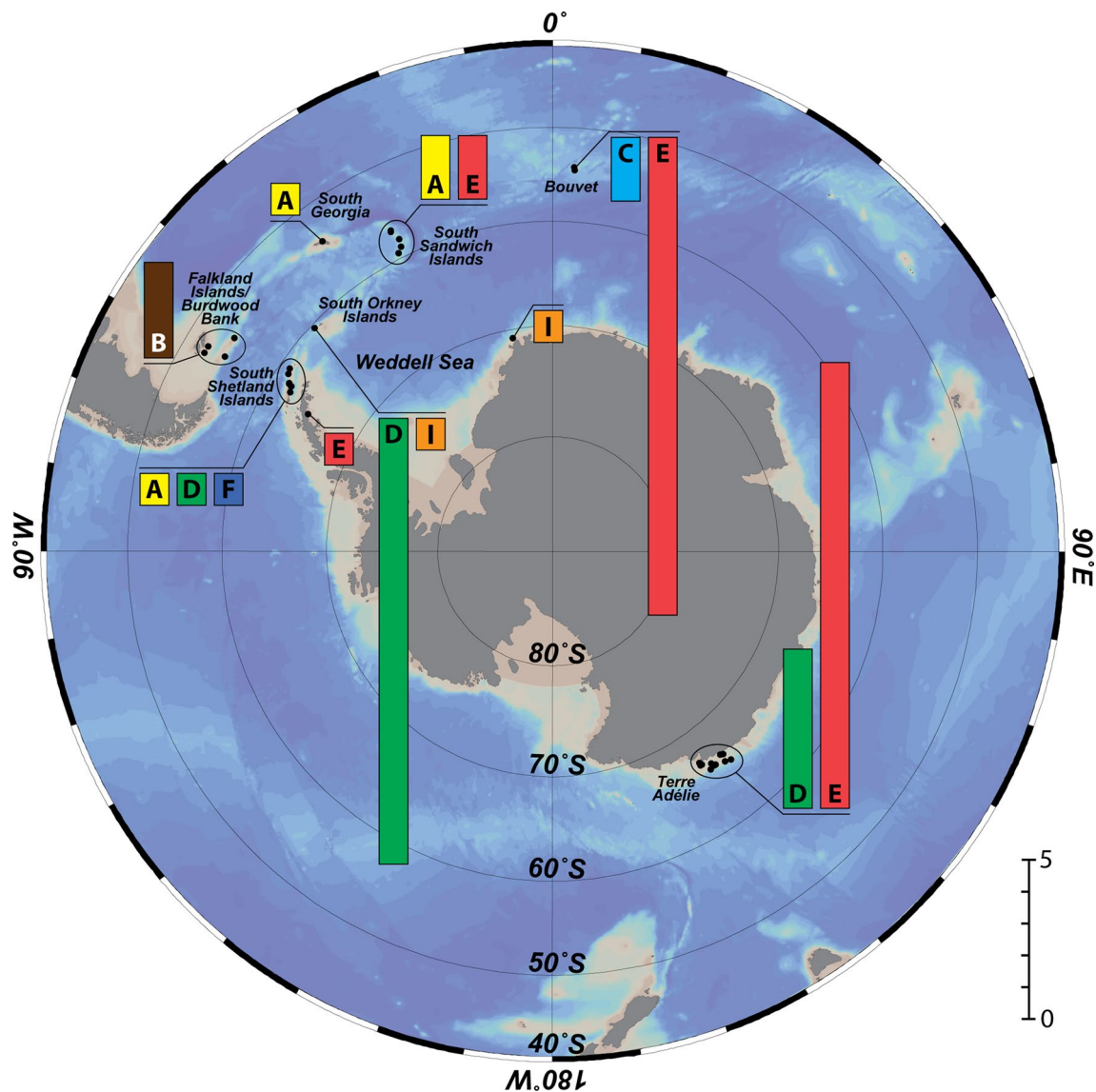
Fig. 1 Sample localities and specimens. Black dots represent sampling sites of Antarctic and sub-Antarctic specimens of the *Colossendeis megalonyx* species complex used for genetic analyses. For each region, each COI clade (identity indicated by the letter) is represented by a single bar. Height of the bar represents the sample size (see the scale at lower right part)

of one individual each of *C. angusta* Sars, 1877 and *C. scotti* Calman, 1915 already used in Dietz et al. (2019) were included as outgroup since that study already showed that those two species were suitable outgroup species to root the phylogeny within *C. megalonyx*. Additionally, a filtered alignment, excluding positions present in less than half of the samples, was generated and another phylogenetic tree was calculated as described above to serve as background information to test for selection.

Further, selection analyses were conducted with HyPhy v. 2.3.13 (Kosakovsky Pond et al. 2005) using default settings as recommended by the authors. Evidence for selection for each codon site was tested with a 'Fast,

Unconstrained Bayesian AppRoximation' (FUBAR; Murrell et al. 2013). FUBAR assumes that the selection pressure is constant along the entire phylogeny. In addition to FUBAR, we also used the 'mixed effects model of evolution' (MEME; Murrell et al. 2012) approach. MEME can detect episodic selection, i.e., sites evolving under positive selection for a subset of the data set. Genes with codons reported under selection by both methods (FUBAR: $pp \geq 0.99$; MEME: $p \leq 0.01$) were further used to identify branches of the phylogenetic tree (rather than sites within the alignment) under positive selection. Therefore, the 'Branch-Site Unrestricted Statistical Test for Episodic Diversification' (BUSTED; Murrell et al.
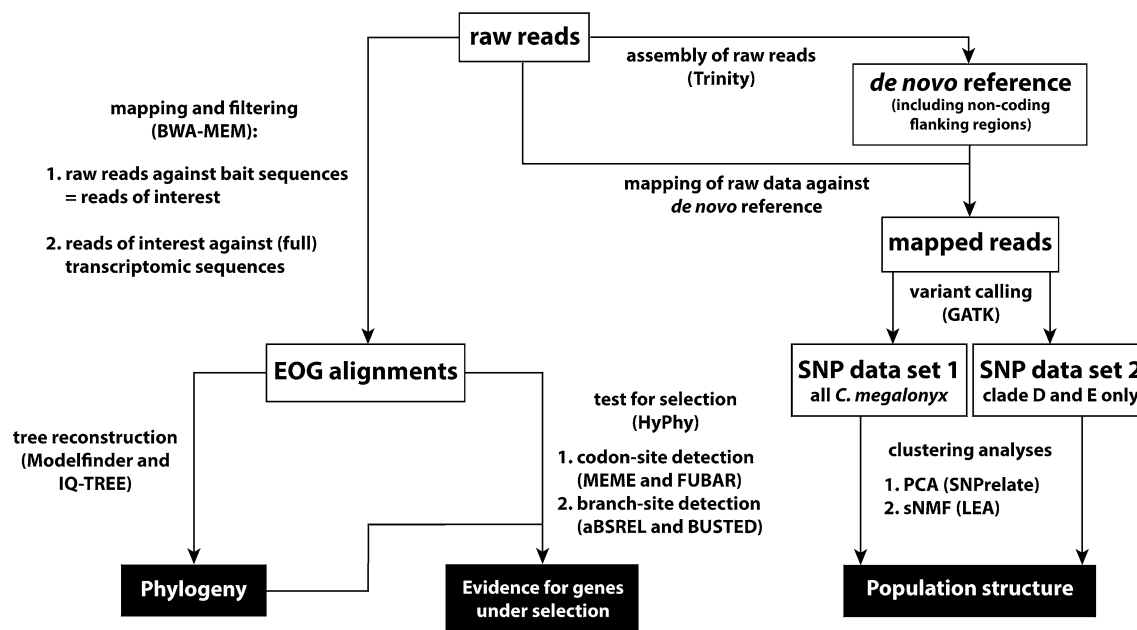
Fig. 2 Flow chart of bioinformatics analyses starting from target enrichment raw reads. Overview of analyses conducted using target hybrid enrichment data generated during this study

2015) at a minimum of one site or branch of a gene was applied using these genes. Finally, the 'Adaptive Branch-Site Random Effects Likelihood' (aBSREL; Smith et al. 2015) model was used to test for each branch whether a subset of sites has evolved under positive selection. Here, both terminal and internal branches were tested.

Genetic GTR distances between individuals were calculated in PAUP* v. 4.0a165 (Swofford 2003).

Gene ontology

Genes potentially under selection reported by FUBAR and MEME were functionally characterized using Gene Ontology (GO) categories. Therefore, the sequence of each gene from the *C. megalonyx* transcriptome was analyzed using Blast2GO (Götz et al. 2008). First, a blastx fast search using the NCBI metazoan database for genes was conducted for gene identification, using an E-value threshold of 10^{-5}. Subsequently, hits were used for annotation and assignment of GO terms for each gene. GO terms were further analyzed with REVIGO (Supek et al. 2011) to reduce and combine closely related gene terms and visualize results. Fisher's exact test was conducted using Blast2GO with a false discovery rate of ≤ 0.05 to test whether GO terms of genes potentially under selection were significantly enriched compared to all genes analyzed.

Single nucleotide polymorphism analyses

Variant calling was conducted as described in Dömel et al. (2019). In brief, a reference assembly based on all raw reads from the samples of *C. megalonyx* was generated with Trinity v. 2.5.1 (Grabherr et al. 2011). Non-coding and more variable flanking regions were included in the data set. Trimmed raw reads of all samples were mapped to the reference assembly using BWA-MEM (Li 2013) and processed with samtools v. 1.6 (Li et al. 2009; Li 2011). Variant calling was conducted with HaplotypeCaller from the GATK v. 4.0.3.0 package (McKenna et al. 2010). Variant calling was performed for two separate data sets: (i) all *C. megalonyx* samples (SNP data set 1, Fig. 2), and (ii) sample from clade D1 and clade E1 only, excluding PS77_257_2_5_2 (SNP data set 2, Fig. 2). The latter was done specifically to obtain SNPs for population genetic analyses, see aim (ii) in the introduction.

To analyze the genetic structure, principal component analyses (PCA) were conducted using the R package SNPRelate v. 1.12.2 (Zheng et al. 2012) with default parameters. For SNP data set 2, plots of the sparse non-negative matrix factorization (sNMF) were calculated to get proportions of the degree of admixture between clades, using the LEA package v. 2.0.0 (Frichot and Francois 2015). *K* values (number of ancestral populations) in the interval 1–20 were tested. The number of repetitions per *K* was set to 40 with 40,000 iterations. The lowest

cross-entropy per *K* value was determined and plotted to choose the most likely *K* value.

Morphology

Morphometric measurements of *C. megalonyx* specimens were carried out using a digital caliper (MarCal IP67, Mahr Metrology, Germany). In total, 133 characters were measured (see Online Resource 1 for a list of characters). It should be mentioned that herein the most recent nomenclature for palps, where the basal article is redefined as the palp's basal process, was used (for details see Cano-Sánchez and López-González 2017).

To test for significant differences of morphological characters between mitochondrial clades, non-parametric unifactorial Kruskal–Wallis *H* tests in combination with Dunn's post hoc tests (Bonferroni corrected) were used in Past v. 3.18 (Hammer et al. 2001).

As the morphometric data set included many missing values, due to articles that broke off during storage, further analyses were based on characters and individuals that had a maximum of 10% missing values. Remaining missing values were imputed using Predictive Mean Matching. Furthermore, only clades with a minimum of three individuals were kept, and clades G, N, and O were excluded from analyses. In addition, relative lengths of all measurements were expressed as a proportion of the trunk length of each specimen. Consequently, two reduced data sets consisting of 92 individuals each were analyzed. Number of characters included was 44 for actual values of measurements (left and right averaged) and 43 for relative lengths. A selection of characters for Linear Discriminant Analysis (LDA) was performed to avoid model overfitting. The heuristic search for the optimal sets of characters was carried out by iteratively using the stepclass function from the R package klaR v.0.6–14 with forward–backwards selection and cross-validation correctness rate as the optimality criterion. The search was organized by picking each of the characters as the starting variable and repeating the procedure ten times. The performance of the character sets was recorded, and the best set was used for a final LDA which was performed using Past.

Morphological distances between mitochondrial clades were calculated as Δ*p* (Safran et al. 2012) using MATLAB R2018b (The MathWorks, Inc., Natick, Massachusetts, United States).

Results

Enrichment and genomic analyses

Sequencing of target hybrid enrichment libraries of specimens of the *Colossendeis megalonyx* species complex resulted in an average number of 426,047 reads per individual (standard deviation (SD) 67,634). All 1607 EOGs were recovered. The average number of recovered genes per individual was 1603.13 genes (99.7%, SD 18%). All genes were found in at least 49 out of 64 individuals (51 when the outgroup species were included). On average, each gene was recovered in 63.75 individuals (99.6%). The shortest gene alignment was 60 bp and the longest 8103 bp.

The length of the concatenated alignment was 1,078,695 bp and it contained 15.9% missing data (ranging from 11.24% to 47.8% between individuals). In the optimal partitioning scheme selected by ModelFinder, all three codon positions were treated as separate partitions. The best model of evolution according to the BIC was GTR + R2 for first and second codon positions and GTR + R3 for third codon positions. The phylogenetic tree (Fig. 3) supported the mitochondrial COI clades identified by Dietz et al. (2015) except for clades D1 and E1. Clade D1 was recovered as paraphyletic concerning most specimens of clade E1. Clade E1 was polyphyletic, as it was divided into two major branches that grouped in different parts of the tree. One branch was represented by one individual from the Eastern Antarctic Peninsula (PS77_257_2_5_2) only. This individual belongs to the nuclear ITS group III, in contrast to the other samples from clade E1 analyzed, which were assigned to ITS group II by Dietz et al. (2015). However, also one individual of clade C was assigned to ITS group II, but according to our data it is clearly separated from the other specimens of this group (clades D1 and E1).

The tree presented herein recovered the larger groupings A + F + I, B + C and D + E, which were consistent with the COI tree (Dietz et al. 2015). Within the specimens belonging to the mitochondrial clades D1 and E1, three geographically restricted clusters per clade were identified. All geographic clusters are supported with bootstrap values of 100%. Nearly all Scotia Arc individuals formed the most basal group of clade D1. Within that group, the single individual from Elephant Island (PN_E010) grouped outside those from the South Orkney Islands (*n* = 13). A single individual from the South Orkneys (YPM48435-9), however, clustered with the individuals from Terre Adélie (*n* = 5). Within clade E1, individuals from Terre Adélie (*n* = 14) were the sister group to those from the South Sandwich Islands (*n* = 2) and Bouvet Island (*n* = 15). Although individuals of both mitochondrial clades D1 and E1 occur in Terre Adélie, they represented two separated groups in the phylogenetic tree. Genetic
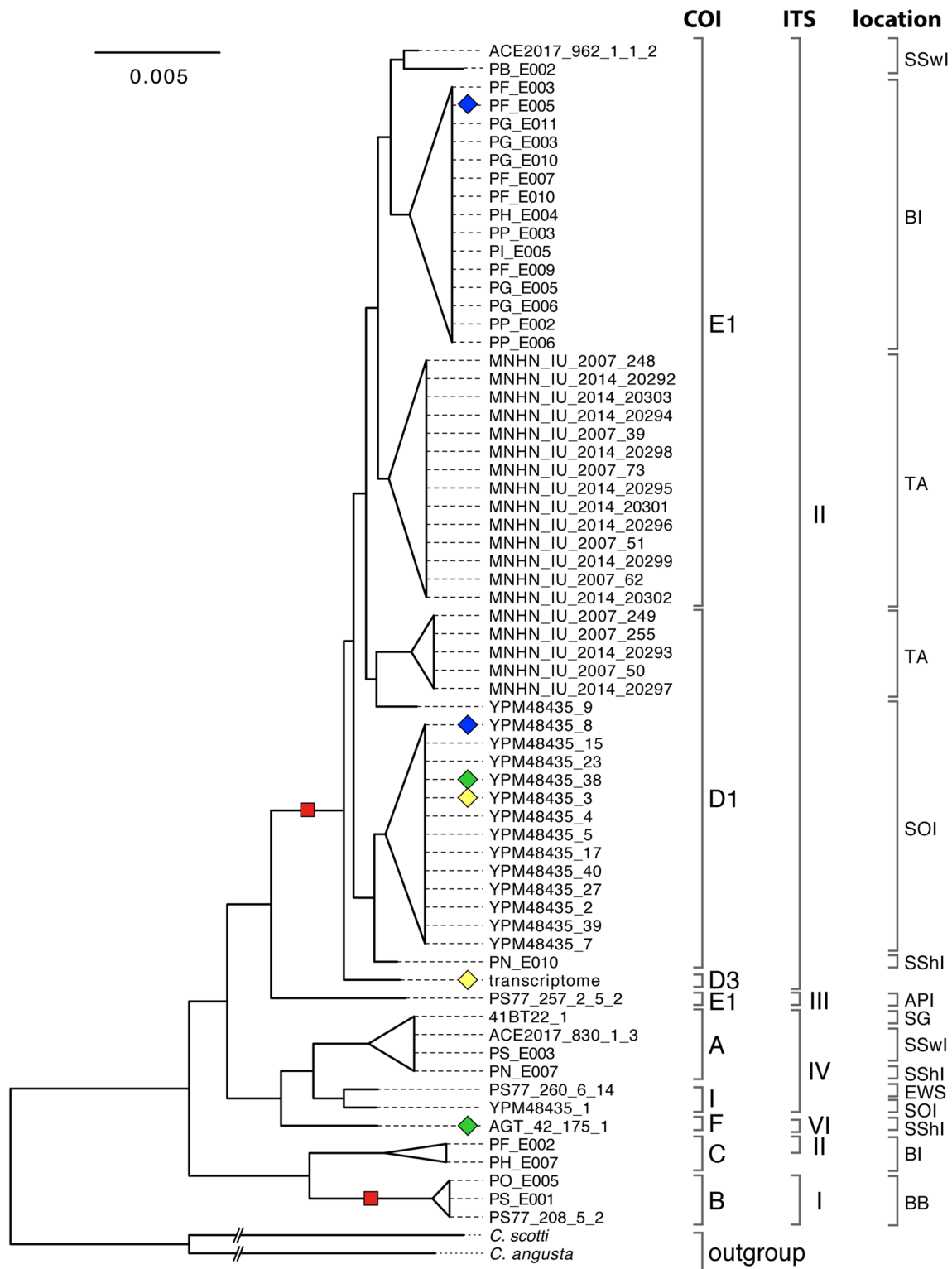
Fig. 3 Maximum-likelihood tree based on concatenated EOG sequences of all genetic samples of the *Colossendeis megalonyx* species complex. Affiliation to COI and ITS groups (Dietz et al. 2015) as well as geographical sampling locations are shown on the right side (*API* Antarctic Peninsula; *BI* Bouvet Island; *BB* Burdwood Bank; *EWS* Eastern Weddell Sea; *SG* South Georgia; *SOI* South Orkney Islands; *SShI* South Shetland Islands; *SSwI* South Sandwich Islands; *TA* Terre Adélie). Nodes or branches for which genes under selection were found with aBSREL and BUSTED are indicated by squares or diamonds, respectively. Each shade represents one gene found to be under selection (squares (red): EOG0910019Q; dark diamonds (blue): EOG0910000D; medium diamonds (green): EOG091000CK; light diamonds (yellow): EOG0910003N)

Table 2 Morphometric (upper right; green) and genomic (lower left; red) distances between lineages of the *Colossendeis megalonyx* species complex represented as heat maps (the darker the greater the distances). *n* is the number of individuals analyzed for each clade and data set

	E (n = 28)	D (n = 15)	A (n = 13)	I (n = 6)	F (n = 8)	C (n = 10)	B (n = 17)
E1 (n = 31)		95.57	107.12	369.95	151.79	72.69	72.21
D (n = 17)	0.0027		149.87	219.12	134.6	72.86	131.62
A (n = 4)	0.0091	0.0091		109.61	213.49	114.63	72.01
I (n = 2)	0.0085	0.0085	0.004		276.43	179.92	118.31
F (n = 1)	0.0092	0.0091	0.006	0.0055		124.19	186.58
C (n = 2)	0.0116	0.0115	0.0119	0.0113	0.0118		94.07
B (n = 3)	0.0126	0.0126	0.0126	0.0121	0.0125	0.0076	
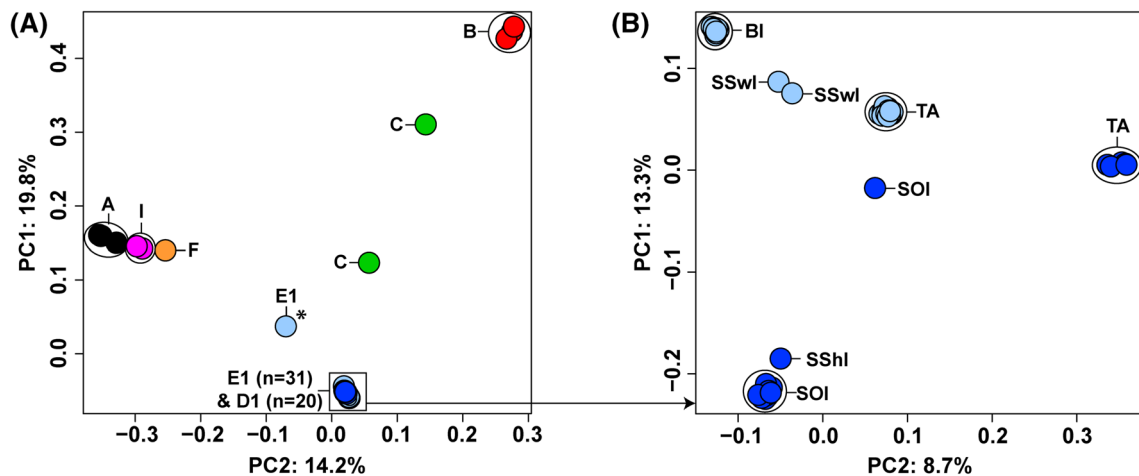
Fig. 4 PCA from genomic data of the *Colossendeis megalonyx* species complex. PCA plots based on genomic data of **a** all samples and **b** samples of the clades D1 (dark) and E1 (light) with the latter excluding specimen PS77_257_2_5_2. PS77_257_2_5_2 was assigned to another ITS group in Dietz et al. (2015) and is indicated with an Asterisk (*) in **a**. *API* Antarctic Peninsula; *BI* Bouvet Island; *BB* Burdwood Bank; *EWS* Eastern Weddell Sea; *SG* South Georgia; *SOI* South Orkney Islands; *SShI* South Shetland Islands; *SSwI* South Sandwich Islands; *TA* Terre Adélie

distances between clades ranged from 0.0027 between clade D and E to 0.0126 between B and E/D/F (Table 2).

Variant calling for SNP data set 1 (all *C. megalonyx*) yielded 13,611 SNPs. The PCA based on those data had 14 significant axes and showed differentiation into five groups, i.e., clades A, B, F, I, and a group consisting of clades D1 and E1 (excluding PS77_257_2_5_2) (Fig. 4a). The remaining individuals, two of clade C and PS77_257_2_5_2 (clade E1), did not cluster within other groups or with each other. For SNP data set 2 (clade D1 and E1 only) 14,904 SNPs were found. PCA of those data revealed six significant axes and showed differentiation into eight groups that mostly show a geographic separation (Fig. 4b). The sNMF analysis supported the results of the SNP-PCA and reported *K* = 4 as the most likely number of clusters (Fig. 5a).

Both mitochondrial clades D1 and E1 were further subdivided into two geographical groups (Fig. 5B). For clade E1, individuals from Terre Adélie and Bouvet Island showed a high proportion of one ancestral population each. Individuals from the South Sandwich Islands showed a high level of admixture of both of those ancestral populations. For clade D1, individuals from the South Orkney Islands and Terre Adélie showed a high proportion of one ancestral population each. One specimen from the South Shetland Islands was mostly assigned to the same ancestral population as those from the South Orkney Islands but also showed a small proportion of ancestry shared with individuals from Terre Adélie of clade E1. One individual from the South Orkney Islands (YPM48435-9) was assigned to similar proportions of all four ancestral populations. This pattern (a similar proportion to all ancestral populations) remained constant with increasing *K*. Only at a value of *K* = 11, the individual was assigned to a separate population, a result clearly rejected by a much higher cross-entropy.
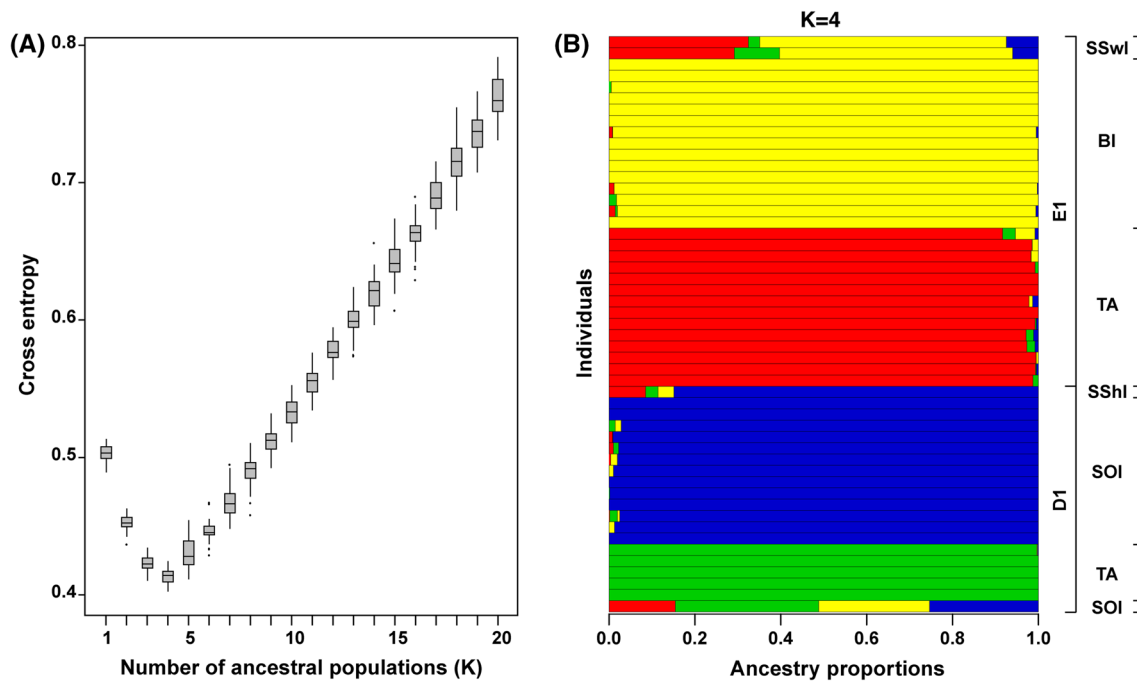
Fig. 5 sNMF analyses of individuals from clades D1 and E1 (excluding specimen PS77_257_2_5_2). **a** Cross-entropy estimates of 1 to 20 ancestral populations (*K* value). **b** Graphical illustration of ancestry proportion estimates with *K* = 4. Estimated ancestry proportions for each specimen are represented by horizontal bars. Clade assignment and sampling location are shown on the right side (*BI* Bouvet; *SOI* South Orkney Islands; *SShI* South Shetland Islands; *SSwI* South Sandwich Islands; *TA* Terre Adélie). The lowest bar represents YPM48435-9

Genes under selection

FUBAR and MEME identified 473 codons in 342 genes and 199 codons in 126 genes, respectively, to be under selection. Of these, 139 codons and 124 genes were shared between the two methods. GO terms were assigned to 93 of the 124 genes under selection. Blast2GO found 265 hits, of which 84 were relevant for biological processes (P), 79 for cellular components (C), and 102 for molecular functions (F). Hits resulted in 119 different GO terms that were further combined to 107 GO terms (P: 39; C: 21; F: 47) by REVIGO (see Online Resources 2 and 3 for figures and tables, respectively).

When testing the 124 shared genes identified with MEME and FUBAR using BUSTED, evidence for selection in 20 genes was supported. Six genes were found to be under selection on 11 branches of the phylogenetic tree using the aBSREL model. Of those, six branches (three genes) corresponded to terminal branches, and five branches were internal ones. Of the latter, only three branches (two genes) were well supported in the phylogenetic tree (pp 100%). Four genes were found in both branch-site tests, including those where terminal branches within different clades were supposed to be under selection and one where two internal nodes were supposed to be under selection (Fig. 3). The four genes found to be under selection in all selection tests were further investigated (Table 3). For three of these genes, GO terms could be determined.

Morphometric analyses

Morphometric measurements were taken for 103 individuals (a table including all measurements is provided in Online Resource 4). After the averaging of bilateral characters, the final data set consisted of 76 characters per specimen. Of those, 68 characters had significant differences between mitochondrial clades (see Online Resource 5 for list of characters and significant *p*-values). The 7th and the 9th (most distal) articles of the palpus and the 5th article of the oviger had the most differences between mitochondrial clades (*n* = 8). The clades between which most differences were found were clades A and B (*n* = 25) and clades C and I (*n* = 24). Morphometric distances between clades ranged from 72.02 between A and B to 369.95 between clade E and I (Table 2).

The best character combination for the absolute lengths consisted of the trunk length, 'WL2 femur' (length of femur of second walking leg) and four palp articles (palp 2, 5, 7, and 9). For the relative lengths the best character combination consisted of 'proboscis thickest' (diameter of proboscis at thickest part) and six palp articles (basal palp and palps 2, 3, 5, 7, and 9) (see Online Resource 6 for detailed result

Table 3 List of genes
potentially under selection

Gene ID	Amino acid position	Description (Blastx-fast)	InterPro GO IDs and names
EOG091000CK	208, 226	Hexosaminidase D-like	0004553 (F: hydrolase activity, hydrolyzing *O*-glycosyl compounds) 0005975 (P: carbohydrate metabolic process) 0015929 (F: hexosaminidase activity)
EOG0910000D	3145, 6556	Nesprin-1 isoform X7	0005515 (F: protein binding) 0005635 (C: nuclear envelope) 0034993 (C: meiotic nuclear membrane microtubule tethering complex) 0051015 (F: actin filament binding) 0090286 (P: cytoskeletal anchoring at nuclear membrane)
EOG0910003N	166	SLIT-ROBO Rho GTPase-activating protein 1-like isoform X5	0005515 (F: protein binding) 0007165 (P: signal transduction)
EOG0910019Q	532	Neuroglian-like isoform X2	No GO term

Details (extracted from Blast2GO) about the four genes detected to be under selection with all methods used

of selection analyses based on morphometric data). Cross-validation confusion rate was higher for relative lengths (0.90) in comparison to one of the absolute lengths (0.81) (Table 4). Misassignment in the cross-validated confusion matrices did not exceed two, except for clade E with nine based on the absolute lengths and seven based on the relative lengths (Table 4). In LDA plots, clades were difficult to separate, except for clade I which was separated from the other clades in all plots for the absolute length that did not include the fourth axis and in the plot for the relative length of the second and third axes (Fig. 6).

Discussion

Species diversity within the species complex *Colossendeis megalonyx*

With our study we confirm to a large degree the proposed phylogeny of the *C. megalonyx* species complex proposed by Dietz et al. (2015). By using genomic target hybrid enrichment data, we add much greater support to branches and yield a more detailed resolution. For mitochondrial clade E1, there was one individual (PS77_257_2_5_2) that did not group with the other individuals assigned to clade E1. However, it was the only individual of this clade belonging to ITS group III. This mito-nuclear discordance has been reported before by Dietz et al. (2015). The ITS sequence of this individual was very similar to those of others found in the same location (Eastern Antarctic Peninsula), which

belonged to the mitochondrial clade N3. The position of clade N3 in the Bayesian mitochondrial tree by Dietz et al. (2015) was similar to that of the discussed individual in the phylogenomic tree (sister to clades D and E). This confirms that the misleading placement of PS77_257_2_5_2 in the mitochondrial tree is a result of introgression of mitochondrial DNA. Another clade with mito-nuclear discordance according to Dietz et al. (2015) was clade C, for which the ITS sequence of only one (PF_E002) of the two analyzed individuals is known. In contrast to other members of clade C, PF_E002 belongs to ITS group II, that also includes clade E1 individuals from the same location. Here our results agree with the mitochondrial data, as clade C is resolved as sister to clade B, suggesting that the ITS sequence in this individual is a result of hybridization.

An additional benefit of the genomic data is the increased resolution of specimens assigned to mitochondrial clades D1 and E1. These were both assigned to ITS group II in Dietz et al. (2015). However, target hybrid enrichment data clearly show that D1 is paraphyletic with respect to E1. One possible explanation for this is that the group ancestrally had the mitochondrial DNA of clade D, and a subgroup of it acquired the mitochondrial DNA of clade E by introgression from an unknown source. In contrast to the ITS data, which suggest that clades D1 and E1 belong to a single species, target hybrid enrichment data detect clear signatures of divergence even in sympatry (Terre Adélie). Trawls are often dragged for a few kilometers and potentially cover multiple habitats. However, the multiple occurrences in one trawl are most likely due to habitat sharing. As the clades D and E

Table 4 Cross-validation confusion matrices for morphometric data sets of the *Colossendeis megalonyx* species complex using absolute and relative lengths

| | Absolute lengths (correctness rate: 0.82) | | | | | | | Relative lengths (correctness rate: 0.90) | | | | | | |
	Clade A	Clade B	Clade C	Clade D	Clade E	Clade F	Clade I	Clade A	Clade B	Clade C	Clade D	Clade E	Clade F	Clade I
Clade A	**11**	0	0	1	1	0	0	**10**	0	1	0	1	0	1
Clade B	1	**14**	0	1	0	0	0	1	**14**	0	1	0	0	0
Clade C	0	0	**10**	0	0	0	0	0	0	**10**	0	0	0	0
Clade D	1	0	2	**11**	0	0	0	1	0	1	**11**	1	0	0
Clade E	3	0	2	1	**19**	0	1	2	0	0	4	**19**	0	1
Clade F	0	0	0	1	1	**5**	0	0	0	0	1	1	**5**	0
Clade I	0	0	0	0	0	0	**6**	0	0	0	0	0	0	**6**

Diagonal elements in the table/the matrix are in bold

were distinguished by both mitochondrial and nuclear data, they are genetically clearly separated and reproductively isolated from each other. Kekkonen and Hebert (2014) stated that reproductively isolated groups that occur in sympatry represent different species (Biological Species Concept). By this criterion, the two groups should be treated as distinct species. Hence, we can assume that the benthic zone of Terre Adélie represents a diverse habitat offering potential for multiple distinct niches and also served as a secondary contact zone where differentiated clades met again after allopatric divergence. Similarly, in the South Orkney Islands, two clearly differentiated lineages belonging to the mitochondrial haplogroup D1 are present according to our data. If these are considered as distinct species, our phylogeny suggests that the other geographically restricted lineages between this group are also separate species, suggesting that the number of species within the *C. megalonyx* complex may have been strongly underestimated with mitochondrial data. As not all clades included in the species complex *C. megalonyx* were analyzed within the present study, no conclusion about the actual number of species within the complex can be made. However, within the subsample used, there are already more lineages than previously detected due to further differentiation within mitochondrial clades, e.g., clade E1.

Differentiation of lineages from the species complex *C. megalonyx* using morphometric data was possible but difficult. This might be explained by high morphological variation within clades. High intraspecific morphological variation has also been reported for Antarctic nematodes (Hauquier et al. 2017). A previous study revealed that there is a high degree of sexual dimorphism within members of the different mitochondrial clades (Spaak 2010). For clades C and E, 30% of all analyzed characters differed significantly between sexes. For clades A, B, and D, this value was much lower (4%) (Spaak 2010). We did not reveal an obvious separation of male and female measurements within the present study, but the increased variability due to sexual dimorphism might represent an issue for PCA based on such morphometric data. However, Dietz et al. (2013) used similar characters and analyses and were able to distinguish *C. tenera* Hilton, 1943 from the species complex *C. megalonyx*. Hence, the shortage of distinct morphological differences is most likely due to the more recent divergence of lineages within this species complex.

Due to the limited sampling of taxa, our results cannot clearly resolve whether *C. megalonyx* originated inside or outside the Antarctic. However, Dietz et al. (2019) recently found that the taxon is nested within a large Antarctic radiation which diversified mostly within the Antarctic, dispersing to other regions multiple times. This supports the scenario that *C. megalonyx* is originally an Antarctic taxon, containing only one clade which dispersed to South
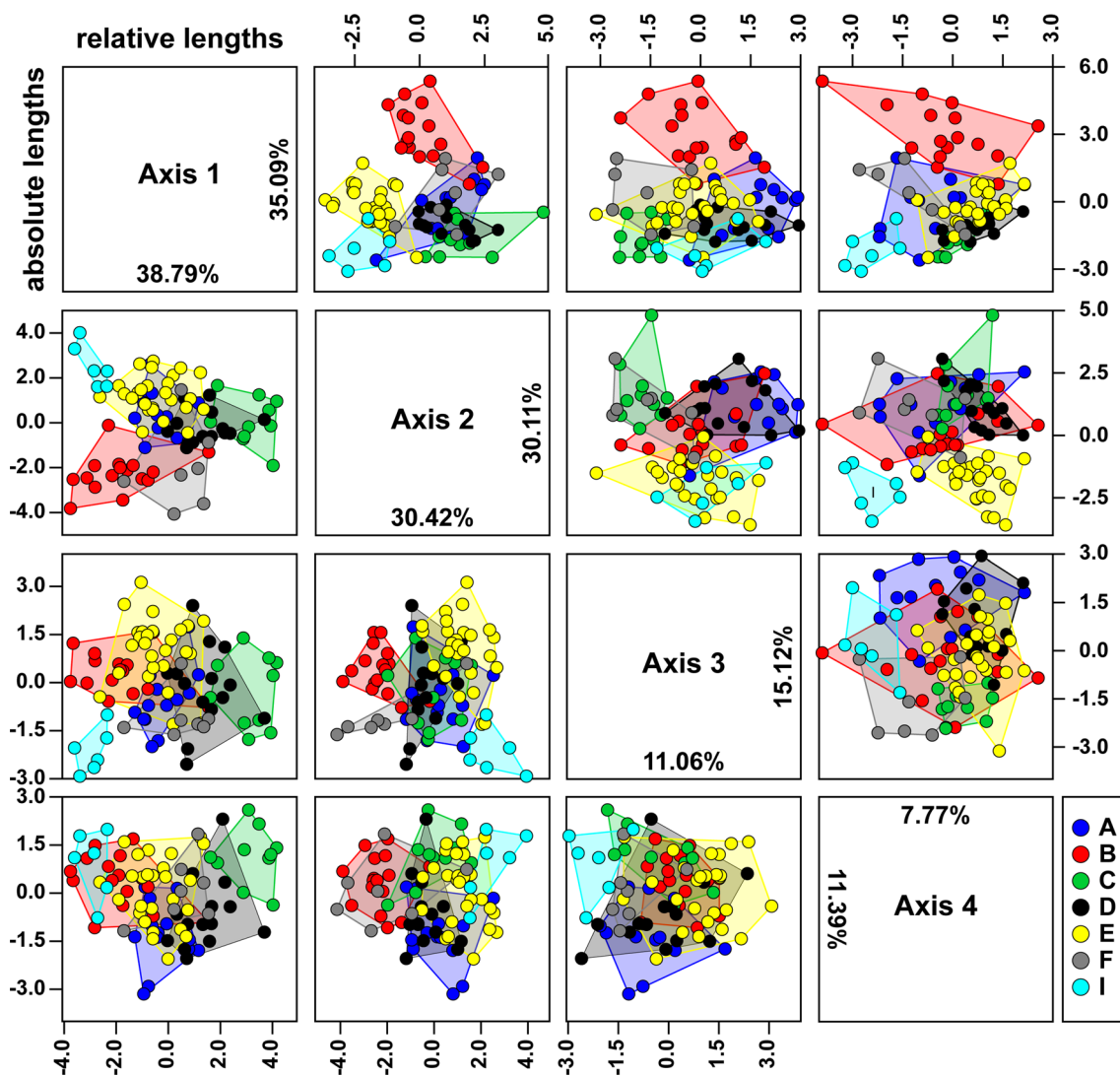
Fig. 6 LDA from morphometric data of the *Colossendeis megalonyx* species complex using the best combination for absolute and relative lengths. LDA plots show all four axes of the same analysis conducted with the reduced morphometric data set for absolute (lower left) and relative (upper right) lengths each. Key for clade assignment on the right

America, as indicated by the mitochondrial tree with a larger sampling within the complex (Dietz et al. 2015).

Intraspecific connectivity within the species complex *Colossendeis megalonyx*

Population genetic analyses using SNP data revealed that intraspecific connectivity among geographically distinct populations is restricted. But even more, our data hints at further divergence within locations. Within clade D1, one individual sampled from the South Orkney Islands, YPM48435_9 is clearly separated from the other 14 specimens of the same mitochondrial clade and location, being closely related to the Terre Adélie specimens of clade D1. In the sNMF analysis, the specimen was assigned to all four ancestral clusters with similar probability each. This can have three different reasons: (i) sample contamination, (ii) hybridization, or (iii) the specimen belongs to a separate population despite the COI sequence being identical to other specimens from the South Orkney Islands. Contamination can be regarded as unlikely as observed heterozygosity of this individual was not higher than that in the other individuals. If DNA was contaminated with DNA from (or more) other specimens, fixed differences between these would have been reported as heterozygous SNPs, thus artificially increasing heterozygosity. This was not the case. The individual could also be a hybrid of D1 and E1 individuals. However, hybridization between four distinct clusters, as suggested by the sNMF results, is unlikely. Furthermore, we would expect higher heterozygosity in a hybrid specimen.

Therefore, we suggest that this specimen may represent an independently evolving lineage, with the identical COI haplotype being the result of mitochondrial introgression. In Dietz et al. (2015), the samples from the South Orkney Islands already showed a high species diversity, and nine different mitochondrial clades occur in sympatry. This makes the presence of several intraspecific populations and introgression more likely. In general, the Scotia Sea has been proposed as a hotspot for speciation especially for brooding species (Allcock and Strugnell 2012) and demonstrated for the bivalve *Lissarca notorcadensis* Melvill and Standen, 1907 (Linse et al. 2007). Nothing is known about the reproductive mode of *Colossendeis*, but the limited connectivity between several populations also within proposed lineages leads to the assumption that more species than previously expected exist within the species complex *C. megalonyx*. In this study, we found that only (nuclear) genomic data can unveil the very shallow divergences and the possible young speciation events in pycnogonids.

Evidence of positive selection in the different lineages of *Colossendeis megalonyx*

For the first time, we explicitly searched for signatures of positive selection in the *C. megalonyx* species complex and reported evidence for this in about 8% of all genes analyzed. Genes identified covered a broad range of biological processes, cellular components, and molecular functions. However, many of these genes had only rather higher-level GO categories, i.e., assignments to specific gene functions were not possible. GO term enrichment did not support the dominance of any particular group of functional pathways in the data set. However, it should be noted that when designing the baits, the number of available genes was already limited to 1607 EOGs. The latter was due to the fact that the bait design was based on single-copy genes of spiders that had to be present in the (most likely incomplete) transcriptome of *C. megalonyx*.

Nevertheless, while it might be difficult to assign specific functions to the genes inferred to be under selection, the fact that the same genes were reported to be under positive selection on different branches suggests that they are candidate genes that were potentially involved in the adaptive radiation into independently evolving lineages. We focused our analyses on branches leading to distinct lineages to test for the role of positive selection rather than interpreting evidence in terminal branches. With this approach still about 0.4% of all analyzed genes were found to be under selection. Those genes included one gene (EOG0910000D) associated with nesprin that is part of the LINC (Linker of Nucleoskeleton and Cytoskeleton) complex and one gene (EOG091000CK) relevant for fundamental processes associated with hexosaminidase which plays a role in the carbohydrate metabolic process. No direct biological relevance of these two genes for speciation processes can be deduced at this stage. Another gene of interest was EOG0910019Q as it was found to be under selection in two branches at a deeper level within the phylogeny (one branch that grouped all individuals from clade D and E1 as well as another branch that grouped all individuals from clade B). A BLAST search linked this gene to neuroglian proteins, associated with neuroglia formation and thus probably relevant for the neural system. The importance of adaptation to the cold in neural systems has been shown for fish. Antarctic fishes adapted to their constantly cold environment and showed a relatively high neuronal conduction velocity also at temperatures below zero degree Celsius (Macdonald 1981; Montgomery and Macdonald 1990). Of course, neural systems in invertebrates differ from those in vertebrates, but similar challenges to adapt to the cold can be assumed. In fact, positive selection in this gene was detected for the branch of clade B, the only clade analyzed that occurs outside the Antarctic convergence (Burdwood Bank and the Falkland Islands) where the water temperature is several degrees higher than south of the polar front. However, an assumption of adaptation due to temperature cannot be made as other clades for which the same gene was inferred to be under selection (D1, D3, and E1) occur only south of the Antarctic convergence, like other analyzed clades in which no evidence for selection could be found. Hence, we refrain from more detailed discussions about functional aspects at this point because without a larger data set (sample size and several more genes analyzed) and functional validation of genes, these observations are first evidence, and it is highly recommended not to start storytelling about underlying principles (Pavlidis et al. 2012). Hence, although the importance of adaptation for speciation remains unclear, it has to be considered as a relevant factor and there is an urgent need to improve the data basis for such studies (see below).

Assuming that adaptive processes took place, differences in morphological traits are expected, too. Indeed, morphological differences between individuals of distinct mitochondrial clades can be detected using morphometric measurements. Characters supposed to be influenced by selection (e.g., proboscis, eyes or claws) did not stand out in terms of numbers of significant differences between clades, but if significant differences were found, they always distinguish only one or two specific clades from the others. For example, significant differences in morphometric measurements of proboscis characters were only found between clade B and other clades. Also, significant differences in the relative claw length were found only between clade B or I and other clades. Finally, significant differences in eye structures were predominantly found between clade A and other clades. However, the clades with which significant differences were

found varied between characters and did not show a clear pattern.

Other characters potentially relevant for selection that should be considered for future studies include the inner structures of the proboscis and smaller structures of other body parts. Wagner et al. (2017) analyzed proboscides of various genera and found remarkable differences between taxa, most likely dependent on prey preferences. Also, setae and pores scattered across the body could be of interest. Setae are found to be expressed in various forms and have sensory functions (Lehmann et al. 2017). The pores represent gland openings (Hess et al. 1996; Lehmann et al. 2017), however, they also act as part of the respiratory system and are responsible for cutaneous gas exchange. Lane et al. (2017) described limitations of body size in pycnogonids due to the need to take up sufficient amounts of oxygen. This limit, however, probably varies between species that occur inside and outside of the Antarctic convergence as well as for eurybathic and stenobathic species due to different amounts of available oxygen.

Comparing the sea spider species complexes of *Colossendeis megalonyx* and *Pallenopsis patagonica*

The *C. megalonyx* species complex was compared with the sea spider species complex *Pallenopsis patagonica* based on the results of single-marker analyses by Dömel et al. (2017). That study revealed that the species complexes, both occuring in the Southern Ocean, show comparable genetic distances for the distinct lineages based on COI. A recent study on the *P. patagonica* species complex with a similar focus as this study assessed genetic divergence by conducting a target hybrid enrichment approach using the same bait set as used herein for the *C. megalonyx* complex (Dömel et al. 2019). In combination with morphological variation, species diversity within the species complex was analyzed and factors leading to recurrent speciation events were assessed. The comparison of the phylogenomic trees of both species complexes underlines the benefit of analyzing thousands of genome-wide markers obtained from target hybrid enrichment in comparison to single-marker (e.g., COI and ITS) approaches with respect to the resolution and support of both shallow and deep nodes. When comparing previous studies there was a notable difference between the species complexes because mito-nuclear discordance was only found for the *C. megalonyx* species complex. Target hybrid enrichment data revealed one case of mito-nuclear discordance in lineages that, however, are closely related to each other for the *P. patagonica* species complex, hence this pattern was explained due to a lack of resolution in the mitochondrial data rather than hybridization as proposed for the *C. megalonyx* species complex. A further difference between the

C. megalonyx and *P. patagonica* species complexes is the evidence for positive selection, which is so far lacking for the latter. This might be based on the fact that only half the genes targeted could be recovered for the *P. patagonica* species complex probably due to the fact that baits were originally designed for the *C. megalonyx* species complex. But also biological differences, e.g., in reproductive strategies, should be considered as a factor for differences in speciation scenarios. The reproductive mode remains unclear for *Colossendeis* but it most likely differs from the brooding strategy of *Pallenopsis* for which egg-carrying males were caught frequently (Hübner et al. 2017). Especially if *Colossendeis* larvae (which have never been reported) were to detach and become free-floating, geographic separation would represent smaller barriers for this taxon than for the brooder, *Pallenopsis*. This would explain the more frequent detection of speciation reversal within the *C. megalonyx* species complexes in comparison to the *P. patagonica* species complex, but not the strong geographic divergence within lineages. Alternatively, pre-zygotic barriers might be less stable among lineages within the *C. megalonyx* species complex, which hybridize with each other at times. In the case that there is no fitness reduction for the hybrids, speciation reversal can occur.

Limitations and recommendations for future research

Limitations of samples from remote habitats

One limitation when performing evolutionary studies in remote areas and inaccessible habitats such as the Antarctic benthic zone is the availability of specimens (Kaiser et al. 2013). Furthermore, the quick processing and correct preservation of samples are essential, especially for genomic or transcriptomic studies. This holds true in particular for sea spiders because often single specimens did not contain DNA of suitable quality and quantity for genomic analyses (especially long read sequencing or de novo genome sequencing). Three attempts for de novo sequencing were unsuccessful as part of this project and thus gene annotation information and bait design options were limited.

Recommendations for analyses of non-model organisms

The findings of our study can advise future research on non-model organisms. Depending on the question, target hybrid enrichment as conducted here has obvious strengths but also limitations. If the aim is to screen for candidate genes under selection as broadly as possible, a better gene coverage than available for *C. megalonyx* should be aimed for. Therefore, different techniques not relying on sparse reference data are recommended, if sufficient funding and

adequate material for analysis is available (transcriptome sequencing, shallow genome resequencing). However, target hybrid enrichment represents a good alternative if that is not the case. A strength of the method is the power to resolve phylogenies. Thus, for studies that aim to resolve species diversity, our results demonstrate the exceptional benefit of target hybrid enrichment data, as even with few specimens, the resolution is much better and more robust than earlier results based on a few genes (Krabbe et al. 2010; Dietz et al. 2015). For the inference of population genetic processes, this study provides a glimpse into the yet largely unrecognized power of SNPs obtained from flanking regions co-enriched with targeted gene sequences. Even with few specimens, as were available here, clear patterns of population structure could be derived. Furthermore, individual outliers could be confidentially identified as such, because the large number of markers provided a greater credibility. This strength has also been reported by a simulation study indicating that the number of markers partly compensates for the number of available specimens (Willing et al. 2012). Still, while we find distinct population patterns, the lack of potential source or stepping stone populations limits the interpretation of the patterns. Thus, while this technique allows for robust and highly resolved data even if only one specimen for a population analyzed, we recommend investing on maximizing sample size per population to derive conclusions about demographic processes.

Conclusion

We generated one of the largest genomic data sets to study species divergence and population patterns within a Southern Ocean benthic species complex. On average, 99.6% of the 1607 targeted genes were recovered per individual for the *Colossendeis megalonyx* species complex using hybrid enrichment. This resulted in an alignment with 1,078,695 bp and 15.9% of missing data. The genomic data supports the previously reported phylogeny, but substantially improves resolution and branch support. Our data demonstrate that genetically distinct species of the *C. megalonyx* species complex co-occur in sympatry at several locations and that gene flow among geographically separated populations is very limited. We identified significant morphological differences between members of the complex and revealed several genes under positive selection. Genes under selection were found on terminal and a few deep branches and add no evidence for adaptive speciation in sympatry. Therefore, our data shows that positive selection likely played a role in shaping gene divergence within the *C. megalonyx* species complex, yet there is rather little support for an ecological speciation. Due to the limited number of specimens available

from different regions, statistical power is, however, limited. Therefore, we encourage subsequent studies to use the integrative genomics and morphometric approach proposed here to test competing scenarios of Southern Ocean macrozoobenthos speciation.

Compliance with ethical standards

References

Abdelkrim J, Aznar-Cormano L, Fedosov AE, Kantor YI, Lozouet P, Phuong MA, Zaharias P, Puillandre N (2018) Exon-capture-based phylogeny and diversification of the venomous gastropods (Neogastropoda, Conoidea). Mol Biol Evol 35(10):2355–2374. https://doi.org/10.1093/molbev/msy144

Allcock AL, Strugnell JM (2012) Southern Ocean diversity: new paradigms from molecular ecology. Trends Ecol Evol 27:520–528. https://doi.org/10.1016/j.tree.2012.05.009

Arango CP, Soler-Membrives A, Miller KJ (2011) Genetic differentiation in the circum-Antarctic sea spider *Nymphon australe* (Pycnogonida; Nymphonidae). Deep Sea Res II 58:212–219. https://doi.org/10.1016/j.dsr2.2010.05.019

Arnaud F, Bamber RN (1987) The biology of Pycnogonida. In: Blaxter JHS, Southward AJ (eds) Advances in marine biology. Academic Press, Cambridge, pp 1–96

Aronesty E (2011) Command-line tools for processing biological sequencing data. ea-utils. https://github.com/ExpressionAnalysis/ea-utils

Aronson RB, Thatje S, Clarke A, Peck LS, Blake DB, Wilga CD, Seibel BA (2007) Climate change and invisibility of the Antarctic benthos. Annu Rev Ecol Evol Suppl 38:129–154

Ballesteros, JA, Setton, EVW, López, CES, Arango, CP, Brenneis, G, Brix, S, Cano-Sánchez, E, Dandouch, M, Dilly, GF, Eleaume, MP, Gainett, G, Gallut, C, McAtee, S, McIntyre, L, Moran, AL, Moran, R, López-González, PJ, Scholtz, G, Williamson, C, Woods, HA, Wheeler, WC, Sharma, PP (2020). Phylogenomic resolution of sea spider diversification through integration of multiple data classes. bioRxiv 2020.01.31.929612. https://doi.org/10.1101/2020.01.31.929612

Bergh R (1884) Report on the Nudibranchiata dredged by H.M.S. Challenger during the years 1873–1876 Rep Sci Res Voyage H.M.S. Challenger. 10:1–154, pls. I-XIV

Brandt A (1991) Colonization of the Antarctic shelf by the Isopoda (Crustacea, Malacostraca). Berichte zur Polarforschung 98:1

Breinholt JW, Earl C, Lemmon AR, Lemmon EM, Xiao L, Kawahara AY (2018) Resolving relationships among the megadiverse butterflies and moths with a novel pipeline for anchored phylogenomics. Syst Biol 67:78–93. https://doi.org/10.1093/sysbio/syx048

Briggs J (2003) Marine centres of origin as evolutionary engines. J Biogeogr 30:1–18

Calman WT (1915) Pycnogonida. British Antarctic ("Terra Nova") Expedition, 1910. Zoology 3:1–74

Cano-Sánchez E, López-González PJ (2017) Basal articulation of the palps and ovigers in Antarctic *Colossendeis* (Pycnogonida; Colossendeidae). Helgol Mar Res. https://doi.org/10.1186/s10152-016-0474-7

Chernomor O, von Haeseler A, Minh BQ (2016) Terrace aware data structure for phylogenomic inference from supermatrices. Syst Biol 65:997–1008. https://doi.org/10.1093/sysbio/syw037

Clarke A (2008) Antarctic marine benthic diversity: patterns and processes. J Exp Mar Biol Ecol 366:48–55

Clarke A, Crame JA (1989) The origin of the Southern Ocean marine fauna. Geol Soc Spec Publ 47:253–268. https://doi.org/10.1144/GSL.SP.1989.047.01.19

Clarke A, Johnston NM (2003) Antarctic marine benthic diversity. Oceanogr Mar Biol Annu Rev 41:47–114

Convey P, Stevens MI, Hodgson DAL, Smellie J, Hillenbrand CD, Barnes DKA, Clarke A, Pugh PJA, Linse K, Cary SC (2009) Exploring biological constraints on the glacial history of Antarctica. Quat Sci Rev 28:3035–3048

Coyne J, Orr H (2004) Speciation. Sinauer Associates, Sunderland

Dieckmann U, Doebeli M, Metz J, Tautz D (2004) Introduktion. In: Dieckmann U, Doebeli M, Metz J, Tautz D (eds) Adaptive speciation. Cambridge University Press, Cambridge

Dietz L, Krapp F, Hendrickx ME, Arango CP, Krabbe K, Spaak JM, Leese F (2013) Evidence from morphological and genetic data confirms that *Colossendeis tenera* Hilton, 1943 (Arthropoda: Pycnogonida), does not belong to the *Colossendeis megalonyx* Hoek, 1881 complex. Org Divers Evol 13:151–162. https://doi.org/10.1007/s13127-012-0120-4

Dietz L, Arango CP, Halanych KM, Harder AM, Held C, Mahon AR, Mayer C, Melzer RR, Rouse G, Weis A, Wilson NG, Leese F (2015) Regional differentiation and extensive hybridisation between mitochondrial clades of the Southern Ocean giant sea spider *Colossendeis megalonyx*. R Soc Open Sci 2:140424

Dietz L, Dömel JS, Leese F, Lehmann T, Melzer RR (2018) Feeding ecology in sea spiders (Arthropoda: Pycnogonida): what do we know? Front Zool 15:7. https://doi.org/10.1186/s12983-018-0250-4

Dietz L, Dömel JS, Leese F, Mahon AR, Mayer C (2019) Phylogenomics of the longitarsal Colossendeidae: the evolutionary history of an Antarctic sea spider radiation. Mol Phylogenet Evol 136:206–214

Dömel JS, Convey P, Leese F (2015) Genetic data support independent glacial refugia and open ocean barriers to dispersal for the Southern Ocean sea spider *Austropallene cornigera* (Möbius, 1902). J Crust Biol 35:480–490. https://doi.org/10.1163/1937240X-00002351

Dömel JS, Melzer RR, Harder AM, Mahon AR, Leese F (2017) Nuclear and mitochondrial gene data support recent radiation within the sea spider species complex *Pallenopsis patagonica*. Front Ecol Evol 4:139. https://doi.org/10.3389/fevo.2016.00139

Dömel JS, Macher T-H, Dietz L, Duncan S, Mayer C, Rozenberg A, Wolcott K, Leese L, Melzer RR (2019) Combining morphological and genomic data to resolve species diversity and study speciation processes of the *Pallenopsis patagonica* (Pycnogonida) species complex. Front Zool 16:36. https://doi.org/10.1186/s12983-019-0316-y

Eastman J, McCune A (2000) Fishes on the Antarctic continental shelf: evolution of a marine species flock? J Fish Biol 57:84

Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. Trends Ecol Evol 29:51–63. https://doi.org/10.1016/j.tree.2013.09.008

Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst Biol 61:717–726

Fraser CI, Nikula R, Ruzzante DE, Waters JM (2012) Poleward bound: biological impacts of Southern Hemisphere glaciation. Trends Ecol Evol 27:462–471

Fraser CI, Terauds A, Smellie J, Convey P, Chown SL (2014) Geothermal activity helps life survive glacial cycles. PNAS 111(15):5634–5639. https://doi.org/10.1073/pnas.1321437111

Frichot E, François O (2015) LEA: An R package for landscape and ecological association studies. Methods Ecol Evol 6:925–929. https://doi.org/10.1111/2041-210X.12382

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 36(10):3420–3435. https://doi.org/10.1093/nar/gkn176

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq

data without a reference genome. Nat Biotechnol 29:644–652. https://doi.org/10.1038/nbt.1883

Griffiths HJ, Arango CP, Munilla T, McInnes SJ (2011) Biodiversity and biogeography of Southern Ocean pycnogonids. Ecography 34:616–627. https://doi.org/10.1111/j.1600-0587.2010.06612.x

Halanych KM, Mahon AR (2018) Challenging dogma concerning biogeographic patterns of Antarctica and the Southern Ocean. Annu Rev Ecol Evol Syst 49:355–378

Hammer Ø, Harper DAT, Ryan PD (2001) PAST: paleontological statistics software package for education and data analysis. Palaeontol Electron 4(1):1.4

Hauquier F, Leliaert F, Rigaux A, Derycke S, Vanreusel A (2017) Distinct genetic differentiation and species diversification within two marine nematodes with different habitat preference in Antarctic sediments. BMC Evol Biol 17:120. https://doi.org/10.1186/s12862-017-0968-1

Held C (2000) Phylogeny and biogeography of serolid isopods (Crustacea, Isopoda, Serolidae) and the use of ribosomal expansion segments in molecular systematics. Mol Phylogenet Evol 15:165–178. https://doi.org/10.1006/mpev.1999.0739

Held C (2003) Molecular evidence for cryptic speciation within the widespread Antarctic crustacean. Ceratoserolis trilobitoides (Crustacea, Isopoda) In: Huiskes AHL, Gieskes WWC, Rozema J, Schorno RML, vander Vies SM, Wolff WJ (eds) Antarctic biology in a global context. Backhuys, Leiden, pp 135–139.

Held C, Wägele J-W (2005) Cryptic speciation in the giant Antarctic isopod Glyptonotus antarcticus (Isopoda; Valvifera: Chaetiliidae). Sci Mar 69:175–181

Hess M, Melzer RR, Smola U (1996) The eyes of a nobody, Anoplodactylus petiolatus (Pantopoda; Anoplodactylidae). Helgoländer Meeresuntersuchungen 50:25–36

Hilton WA (1943) Pycnogonida from the Pacific. Family Colossendeidae J Entomol Zool Stud 35(1):2–4

Hoek PPC (1881) Report on the Pycnogonida dredged by HMS challenger 1873–76. Rep Sci Results Explor Voyag HMS Chall 3:1–167

Hübner J, Wagner P, Lehmann T, Melzer RR (2017) Testing species delimitation with larval morphology: scanning electron microscopy analysis of protonymphon larvae of two closely related sea spiders, Pallenopsis patagonica (Hoek) and Pallenopsis yepayekae Weis. Invertebr Syst 31:363–374. https://doi.org/10.1071/IS16050

Hugall AF, O'Hara TD, Hunjan S, Nilsen R, Moussalli A (2015) An exon-capture system for the entire class Ophiuroidea. Mol Biol Evol 33:281–294. https://doi.org/10.1093/molbev/msv216

Johannesson K, Butlin RK, Panova M, Westram AM (2017) Mechanisms of Adaptive Divergence and Speciation in Littorina saxatilis: integrating knowledge from ecology and genetics with new data emerging from genomic studies. In: Oleksiak MF, Rajora OP (eds) Population genomics. Springer, Cham, pp 277–301

Kaiser S, Brandão S, Brix S, Barnes DKA, Bowden DA, Ingels J et al (2013) Patterns, processes and vulnerability of Southern Ocean benthos: a decadal leap in knowledge and understanding. Mar Biol 160:2295–2317. https://doi.org/10.1007/s00227-013-2232-6

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589

Kekkonen M, Hebert PD (2014) DNA barcode-based delineation of putative species: efficient start for taxonomic workflows. Mol Ecol Resour 14:706–715. https://doi.org/10.1111/1755-0998.12233

Knox G, Lowry J (1977) A comparison between the benthos of the Southern Ocean and the North Polar Ocean with special reference to the Amphipoda and the Polychaeta. In: Dunbar MJ (ed) Polar Oceans. Arctic Institute of North America, Calgary, pp 423–462

Kocher T (2004) Adaptive evolution and explosive speciation: the cichlid fish model. Nat Rev Genet 5:288–298

Kosakovsky Pond SL, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676–679

Krabbe K (2010) Populationsgenetik und Phylogeographie von Colossendeis megalonyx. Diploma thesis, Ruhr University Bochum.

Krabbe K, Leese F, Mayer C, Tollrian R, Held C (2010) Cryptic mitochondrial lineages in the widespread pycnogonid Colossendeis megalonyx Hoek, 1881 from Antarctic and Subantarctic waters. Polar Biol 33:281–292

Lane SJ, Shishido CM, Moran AL, Tobalske BW, Arango CP, Woods HA (2017) Upper limits to body size imposed by respiratory–structural trade-offs in Antarctic pycnogonids. Proc R Soc Lond [Biol] 284:20171779. https://doi.org/10.1098/rspb.2017.1779

Lehmann T, Heß M, Melzer RR (2017) Sense organs in Pycnogonida: a review. Acta Zool 99:211–230

Lemer S, Kawauchi GY, Andrade SC, González VL, Boyle MJ, Giribet G (2015) Re-evaluating the phylogeny of Sipuncula through transcriptomics. Mol Phylogenetics Evol 83:174–183

Lemmon AR, Emme SA, Lemmon EM (2012) Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst Biol 61:727–744. https://doi.org/10.1093/sysbio/sys049

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio]

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Linse K, Cope T, Lörz A-N, Sands C (2007) Is the Scotia Sea a centre of Antarctic marine diversification? Some evidence of cryptic speciation in the circum-Antarctic bivalve Lissarca notorcadensis (Arcoidea: Philobryidae). Polar Biol 30:1059–1068. https://doi.org/10.1007/s00300-007-0265-3

Macdonald JA (1981) Temperature compensation in the peripheral nervous system: antarctic vs temperate poikilotherms. J Comp Physiol 142:411–418

Mahon A, Arango C, Halanych K (2008) Genetic diversity of Nymphon (Arthropoda: Pycnogonida: Nymphonidae) along the Antarctic Peninsula with a focus on Nymphon australe Hodgson, 1902. Mar Biol 155:315–323

Mayer C, Sann M, Donath A, Meixner M, Podsiadlowski L, Peters RS, Petersen M, Meusemann K, Liere K, Wägele J-W, Misof B, Bleidorn C, Ohl M, Niehuis O (2016) BaitFisher: a software package for multispecies target DNA enrichment probe design. Mol Biol Evol 33:1875–1886. https://doi.org/10.1093/molbev/msw056

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A GK, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. https://doi.org/10.1101/gr.107524.110

Montgomery JC, Macdonald JA (1990) Effects of temperature on nervous system: implications for behavioral performance. Am

J Physiol Regul Integr Comp Physiol 259:R191–R196. https://doi.org/10.1152/ajpregu.1990.259.2.R191

Moran AL, Woods HA, Shishido CM, Lane SJ, Tobalske BW (2018) Predatory behavior of giant Antarctic sea spiders (*Colossendeis*) in nearshore environments. Invert Biol 137:116–123

Morin RD, Bainbridge M, Fejes A, Hirst M, Krzywinski M, Pugh TJ, McDonald H, Varhol R, Jones SJM, Marra MA (2008) Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. Biotechniques 45:81–94. https://doi.org/10.2144/000112900

Munilla T, Soler-Membrives A (2009) Check-list of the pycnogonids from Antarctic and sub-Antarctic waters: zoogeographic implications. Antarct Sci 21:99–111

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL (2012) Detecting individual sites subject to episodic diversifying selection. PLoS Genet 8:e1002764. https://doi.org/10.1371/journal.pgen.1002764

Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K (2013) FUBAR: a fast, unconstrained Bayesian AppRoximation for inferring selection. Mol Biol Evol 30:1196–1205. https://doi.org/10.1093/molbev/mst030

Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin D, Smith DM, Scheffler K, Kosakovsky Pond SL (2015) Gene-wide identification of episodic selection. Mol Biol Evol 32:1365–1371

Nater A, Burri R, Kawakami T, Smeds L, Ellegren H (2015) Resolving evolutionary relationships in closely related species with whole-genome sequencing data. Syst Biol 64:1000–1017. https://doi.org/10.1093/sysbio/syv045

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274

O´Hara TD, Hugall AE, Woolley SNC, Bribiesca-Contreras G, Bax NJ (2019) Contrasting processes drive ophiuroid phylodiversity across shallow and deep seafloors. Nature 565:636–639. https://doi.org/10.1038/s41586-019-0886-z

Pavlidis P, Jensen JD, Stephan W, Stamatakis A (2012) A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. Mol Biol Evol 29:3237–3248. https://doi.org/10.1093/molbev/mss136

Poulin E, Palma A, Feral J-P (2002) Evolutionary versus ecological success in Antarctic benthic invertebrates. Trends Ecol Evol 17:218–222

Rutschmann S, Matschiner M, Damerau M, Muschick M, Lehmann M, Hanel R, Salzburger W (2011) Parallel ecological diversification in Antarctic notothenioid fishes as evidence for adaptive radiation. Mol Ecol 20:4707–4721

Safran R, Flaxman S, Kopp M, Irwin DA, Briggs D, Evans MR, Funk WC, Gray DA, Hebets EA, Seddon N, Scordato E, Symes LB, Tobias JA, Toews DPL, Uy JAC (2012) A robust new metric of phenotypic distance to estimate and compare multiple trait differences among populations. Curr Zool 58:426–439. https://doi.org/10.1093/czoolo/58.3.426

Sars GO (1877) Prodromus descriptionis crustaceorum et pycnogonidarum, quae in expeditione norvegica anno 1876, observavit. Archiv Mathematik og Naturvidenskab 2:337–368

Schluter D (2000) Ecological character displacement in adaptive radiation. Am Nat 156:S4–S16

Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SK (2015) Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol 32:1342–1353

Spaak J (2010) Integrative Taxonomie am Beispiel des *Colossendeis megalonyx* Hoek 1881 Artkomplexes (Chelicerata; Pycnogonida; Colossendeidae). MSc thesis, Ruhr University Bochum

Strugnell J, Watts P, Smith P, Allcock A (2012) Persistent genetic signatures of historic climatic events in an Antarctic octopus. Mol Ecol 21:2775–2787

Sunnucks P, Hales DF (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). Mol Biol Evol 13:510–524. https://doi.org/10.1093/oxfordjournals.molbev.a025612

Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO summarizes and visualizes long lists of Gene Ontology terms. PLoS ONE 6:e21800. https://doi.org/10.1371/journal.pone.0021800

Swofford DL (2003) PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland

Teasdale LC, Köhler F, Murray KD, O'Hara T, Moussalli A (2016) Identification and qualification of 500 nuclear, single-copy, orthologous genes for the Eupulmonata (Gastropoda) using transcriptome sequencing and exon capture. Mol Ecol Resour 16:1107–1123. https://doi.org/10.1111/1755-0998.12552

Thatje S, Hillenbrand CD, Larter R (2005) On the origin of Antarctic marine benthic community structure. Trends Ecol Evol 20:534–540. https://doi.org/10.1016/j.tree.2005.07.010

Thatje S, Hillenbrand CD, Mackensen A, Larter R (2008) Life hung by a thread: endurance of Antarctic fauna in glacial periods. Ecology 89:682–692

Wagner P, Dömel JS, Hofmann M, Hübner J, Leese F, Melzer RR (2017) Comparative study of bisected proboscides of Pycnogonida. Org Divers Evol 17:121–135. https://doi.org/10.1007/s13127-016-0310-6

Weis A, Melzer RR (2012) How did sea spiders recolonize the Chilean fjords after glaciation? DNA barcoding of Pycnogonida, with remarks on phylogeography of Achelia assimilis (Haswell, 1885). Syst Biodivers 10:361–374

Weis A, Meyer R, Dietz L, Dömel JS, Leese F, Melzer RR (2014) *Pallenopsis patagonica* (Hoek, 1881)—a species complex revealed by morphology and DNA barcoding, with description of a new species of *Pallenopsis* Wilson, 1881. Zool J Linn Soc 170:110–131. https://doi.org/10.1111/zoj.12097

Weiss M, Leese F (2016) Widely distributed and regionally isolated! Drivers of genetic structure in *Gammarus fossarum* in a human-impacted landscape. BMC Evol Biol 16:153. https://doi.org/10.1186/s12862-016-0723-z

Weiss M, Weigand H, Weigand AM, Leese F (2018) Genome-wide single-nucleotide polymorphism data reveal cryptic species within cryptic freshwater snail species-The case of the *Ancylus fluviatilis* species complex. Ecol Evol 8:1063–1072. https://doi.org/10.1002/ece3.3706

Willing EM, Dreyer C, van Oosterhout C (2012) Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many sNP markers. PLoS ONE 7:e42649. https://doi.org/10.1371/journal.pone.0042649

Wilson NG, Hunter R, Lockhart S, Halanych K (2007) Multiple lineages and absence of panmixia in the "circumpolar" crinoid *Promachocrinus kerguelensis* from the Atlantic sector of Antarctica. Mar Biol 152:895–904

Wilson NG, Schrödl M, Halanych KM (2009) Ocean barriers and glaciation: evidence for explosive radiation of mitochondrial lineages in the Antarctic sea slug *Doris kerguelenensis* (Mollusca, Nudibranchia). Mol Ecol 18:965–984

Wilson NG, Maschek JA, Baker BJ (2013) A species flock driven by predation? Secondary metabolites support diversification of slugs

in Antarctica. PLoS ONE 8:e80277. https://doi.org/10.1371/journal.pone.0080277

Wood HM, González VL, Lloyd M, Coddington J, Scharff N (2018) Next-generation museum genomics: phylogenetic relationships among palpimanid spiders using sequence capture techniques (Araneae: Palpimanoidea). Mol Phylogenetics Evol 127:907–918. https://doi.org/10.1016/j.ympev.2018.06.038

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328. https://doi.org/10.1093/bioinformatics/bts610