

南极磷虾(*Euphausia superba*)线粒体基因组特征及其分子标记应用*

申欣^{1,3} 孙松^{1,3} 王海青^{1,3} 王敏晓^{1,3} 任建峰^{1,3}
张光涛¹ 刘斌^{1,2,3}

(1.中国科学院海洋研究所 青岛 266071; 2.中国科学院北京基因组研究所 北京 100029;
3.中国科学院研究生院 北京 100039)

提要 为揭示南极磷虾(*Euphausia superba*)(甲壳动物:软甲纲:磷虾目)线粒体基因组特征,及通过比较不同海域南极磷虾的线粒体基因组探讨潜在的分子标记,采用 Long PCR 扩增技术获得来自普里兹湾(Prydz Bay)南极磷虾线粒体基因组 DNA,综合 Primers-walking 和 Shotgun 方法进行测序;通过生物信息学软件(DOGMA、tRNAscan-SE 1.21 和 DnaSP 4.10.7 等)进行基因预测及序列分析。南极磷虾线粒体基因组全长 15 498 bp 以上(部分非编码区未测定),包含 13 个蛋白编码基因、2 个核糖体 RNA 基因和 23 个转运 RNA 基因。与后生动物线粒体基因组标准的基因组成相比,存在 1 个 *trnN* 基因的重复。南极磷虾线粒体基因组的基因排列与泛甲壳动物线粒体基因组的原始排列相比,出现 4 个转运 RNA 基因的易位:*trnL*₁、*trnL*₂、*trnW* 和 *trnI* 以及 1 个 *trnN* 的重复。比较来自普里兹湾(Prydz Bay)和威德尔海(Weddell Sea)南极磷虾的线粒体基因组,作者发现在南极磷虾线粒体基因组的主编码基因中,变异最大的是 *nad2* 基因,差异位点高达 61 个,其次是 *nad5* 基因,差异位点有 12 个;而 *cox1* 基因的变异位点仅有 3 个。*Nad2* 基因和 *nad5* 基因可作为 *cox1* 基因辅助的分子标记,用于分析南极磷虾不同地理种群之间的遗传多样性,为合理利用南极磷虾生物资源提供保障。

关键词 南极磷虾, 线粒体基因组, 海洋生态, 浮游动物, 分子标记

中图分类号 S968.3

磷虾目(Euphausiacea)隶属于甲壳动物亚门(Crustacea)、软甲纲(Malacostraca),是一类海洋大型浮游甲壳动物。由于具有分布广、数量大等特点,磷虾类在海洋浮游生物中占有重要位置,不仅是众多经济鱼类和须鲸类的主要饵料,在海洋生态系统中发挥着重要作用,而且还直接可以作为渔业捕捞的对象,用于食用或者提取生物活性物质等(孙松等, 2001; 孙松, 2002)。

南极磷虾(*Euphausia superba*)是地球上单种生物资源最大的物种之一,其现存量的最新估计值为 6.5—10 亿 t,形成巨大的潜在渔业资源,每年的可捕

捞量高达 1 亿 t,相当于目前全世界每年鱼类和其他甲壳类渔获量的总和(0.99 亿 t),被认为是地球上最大、也是最后一个动物蛋白库,对于人口膨胀、食物短缺的当今世界,南极磷虾资源有着无限的应用前景(孙松, 2002)。与此同时,南极磷虾是南大洋海洋生物食物链的中心,是维持南大洋海洋生态平衡的关键(孙松, 2002; Zhou *et al*, 2004)。在南大洋生物的食物链中,如果南极磷虾灭绝或者大幅度减少,则捕食磷虾的须鲸类和鱼类等也将灭绝或减少,南大洋的生态平衡随之破坏。不幸的是,南极磷虾的丰度变化非常大,在某些情况下南极磷虾较低的丰度会

* 中国科学院百人计划项目: 052417101Q。申欣, 博士, E-mail: shenthin@163.com

通讯作者: 刘斌, 研究员, 博士生导师, E-mail: bliu@ms.qdio.ac.cn

收稿日期: 2007-03-04, 收修改稿日期: 2008-04-08

导致企鹅和信天翁雏鸟的大量死亡(Machida *et al.*, 2004)。

如何在高效开发利用南极磷虾生物资源的同时, 保持南大洋生态系统的平衡? 这需亟待加强对南极磷虾遗传信息的认识。由于线粒体 DNA 序列存在较高的进化速率, 在确定南极磷虾资源管理单元时格外有效。为了更加有效地利用线粒体 DNA 作为遗传标记, 线粒体基因组的信息是十分关键的。Machida 等(2004)测定了采自威德尔海 Weddell Sea(60°25'S, 46°12'W; 图 1)南极磷虾的部分线粒体基因组, 初步揭示了南极磷虾(Weddell Sea)线粒体基因组的基因组特征; 但由于缺乏部分 *srRNA* 基因和 3 个以上 tRNA 基因的信息, 深层次的基因组特征分析以及南极磷虾线粒体基因组主编码基因的变异度比较等方面没有得到探讨。

为了加强对南极磷虾线粒体基因组特征的认识, 比较不同地理群体线粒体基因组主编码基因的变异程度, 从而为确定有效的分子标记提供保障。本文中, 作者报道了采自南极普里兹湾(Prydz Bay)南极磷虾的线粒体基因组序列, 确定了南极磷虾线粒体基因组编码的所有基因(共包含 13 个蛋白质编码基因、2 个核糖体 RNA 基因和 23 个转运 RNA 基因)。与 Machida 等(2004)测定的南极磷虾(Weddell Sea)相比, 完成了 *srRNA* 基因和增加了 4 个 tRNA 基因。本研究还分析了南极磷虾线粒体基因组的碱基组成、基因排列和密码子使用偏好等基因组特征, 并通过两个南极磷虾线粒体基因组的比较分析, 探讨了可用于南极磷虾生态研究的潜在分子标记。本研究将为南极磷虾的群体遗传学、分子生态学和生物资源的合理利用等方面提供有价值的遗传信息。

1 材料和方法

1.1 样品采集和基因组 DNA 的提取

南极磷虾采自南极普里兹湾(Prydz Bay)(64°44'S, 73°00'E; 图 1)。新鲜样品采集后用无水乙醇固定带回实验室。将样品去壳, 用纯水清洗 3—5 次, 室温晾干, 置于 1.5 ml Eppendorf 管中, 加入 550 μ l SDS 裂解液(Tris-HCl 0.1 mM/L, EDTA 0.02 mM/L, SDS 0.025%), 用手术剪剪碎, 加入 2 μ l 蛋白酶 K(20 mg/ml), 小心混匀, 55°C 水浴 2—4 h, 每隔半小时轻摇一次, 至消化液澄清透明为止。等体积酚仿(苯酚:

氯仿: 异戊醇=25:24:1)抽提 2—3 遍, 然后离心 10 min(13000 r/min), 取上清液, 加入 70%体积异丙醇, 置于 -20°C 沉淀基因组 DNA, 70%乙醇洗涤两次, 1 \times TE(含 Rnase)重悬基因组 DNA。通过琼脂糖凝胶电泳检测基因组 DNA 的质量。

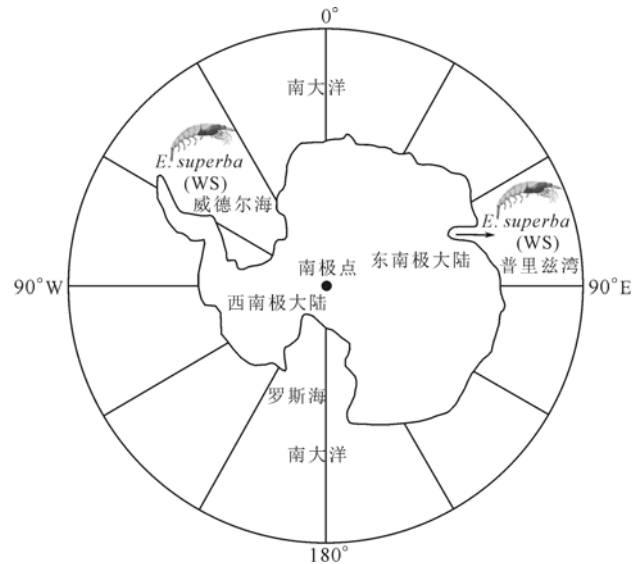


图 1 南极磷虾(Prydz Bay)和南极磷虾(Weddell Sea)的采集位置

Fig.1 The sampling location of *Euphausia superba* (Prydz Bay) and *E. superba* (Weddell Sea)

1.2 Long PCR 扩增及综合 Primers-walking 和 Shotgun 法测序

根据 GenBank 公布的南极磷虾(Weddell Sea)线粒体基因组部分序列(AB084378), 在基因组的五个位置(*cox1*、*cox3*、*nad5*、*cob* 和 *srRNA*)设计正向和反向引物(表 1), 然后根据这些引物在线粒体基因组上的相对位置进行组合配对: *Eus-cox1F/Eus-cox3R*, *Eus-cox3F/Eus-nad5R*, *Eus-nad5F/Eus-cobR*, *Eus-cobF/Eus-srRNAR* 和 *Eus-srRNAF/Eus-cox1R*。Long-PCR 反应体系和反应条件见表 2。PCR 反应产物经 Montage PCR Cleanup Kit(Millipore)纯化后, 使用 ABI 3730x1 DNA Analyzer 进行步移测序。

Long-PCR 反应产物, 经 Hydroshear 仪(Gene Machines, San Carlos, CA)破碎至 2—3 kb 后构建文库。1.5 \times 覆盖度测序, 余下的间隙设计引物进行补测。所有 PCR 反应产物都经 Montage PCR Cleanup Kit(Millipore)纯化后, 使用 ABI 3730x1 DNA Analyzer 进行测序。

表 1 用于扩增南极磷虾(Prydz Bay)线粒体基因组的引物
Tab.1 Primers used in the amplification of *Euphausia superba* (Prydz Bay) mitochondrial genome

引物名称	序列(5→3)
Eus-cox1F	GGTGCATGAGCTGGAATAGT
Eus-cox1R	TTAAGTTGTGCACCGTGAAG
Eus-cox3F	GCACACGGATTTCACACATA
Eus-cox3R	GCTGGCTGAAAAGTGACAAC
Eus-nad5F	TTATGAATTACAGCCCCAGC
Eus-nad5R	AGGTTGAGATGGGTTAGGGT
Eus-cobF	ATCGCAAATAGAGCACTGGT
Eus-cobR	AAAATAATGGTGAATGGGA
Eus-srRNAF	TAAGAATGAGAGCGACGGG
Eus-srRNAR	TTTGGCGGTGTCTTAGTCTAG

表 2 Long-PCR 反应体系和反应条件
Tab.2 Reaction system and condition of the long-PCR

Long-PCR 反应体系		Long-PCR 反应条件		
10×Buffer	3.0 μl	Step	温度	时间
10 mM dNTP	0.6 μl	1	94℃	2 min
5pM primerF	2.0 μl	2	94℃	20 s
5pM primerR	2.0 μl	3	52℃	60 s
模板	1.0 μl	4	65℃	16.0 min
Taq 酶(1U/μl)	1.0 μl	Step2—4 33 cycles		
ddH ₂ O	至 30.0 μl	5	72℃	10 min

1.3 线粒体基因组序列分析及基因预测

测序得到的峰图使用 phredPhrap 进行拼接 (Ewing *et al.*, 1998a,b), 并通过 CONSED 手工检查以避免错拼 (Gordon *et al.*, 1998)。13 个蛋白质编码基因和 2 个核糖体 RNA 的位置首先使用 DOGMA (Wyman *et al.*, 2004; 参数默认) 进行预测。绝大多数的 tRNA 基因和它们的二级结构特征是通过 tRNAscan-SE 1.21 (Lowe *et al.*, 1997) 进行预测, 余下的 tRNA 基因通过类似 tRNA 的二级结构和反密码子进行确定。南极磷虾 (Prydz Bay) 线粒体基因组的基因图谱通过 OGDRAW (Lohse *et al.*, 2007) 进行绘制。通过 DnaSP 4.10.7 (Rozas *et al.*, 2003) 对南极磷虾 (Prydz Bay) 线粒体基因组的密码子使用偏好进行计算。

2 结果

2.1 基因组特征

南极磷虾 (Prydz Bay) 的线粒体基因组长度为 15,498 bp 以上 (部分非编码区未测定), 与 Machida 等 (2004) 测定的南极磷虾 (Weddell Sea) 线粒体基因组相比, 延伸了 892 bp, 完成了 *srRNA* 基因和增加了 4 个

转运 RNA 基因。南极磷虾 (Prydz Bay) 的线粒体基因组编码 38 个基因 (13 个蛋白质编码基因, 23 个转运 RNA 和 2 个核糖体 RNA), 其基因组和后生动物线粒体基因组典型的基因组组成相比, 存在 1 个 *trnN* 基因的重复 (图 2, 表 3)。南极磷虾 (Prydz Bay) 主编码链的 A + T 含量为 68.1% (表 4), 和软甲纲线粒体基因组的平均水平相比差别不大 (Hickerson *et al.*, 2000; Wilson *et al.*, 2000; Yamauchi *et al.*, 2002, 2003; Miller *et al.*, 2004; Yamauchi *et al.*, 2004; Miller *et al.*, 2005; Place *et al.*, 2005; Segawa *et al.*, 2005; Sun *et al.*, 2005; Kilpert *et al.*, 2006; Miller *et al.*, 2006; Ivey *et al.*, 2007; Shen *et al.*, 2007)。南极磷虾 (Prydz Bay) 线粒体基因组提交到 GenBank, 接收号为 EU583500 (图 2)。

2.2 蛋白质编码基因

南极磷虾 (Prydz Bay) 线粒体基因组蛋白质编码基因的精确位置是通过 DOGMA 及与近缘物种线粒体基因的 BLAST 比对后获得的。13 个蛋白质编码基因中, 有 9 个基因在重链上编码 (*atp6*、*atp8*、*cox1*、*cox2*、*cox3*、*cob*、*nad2*、*nad3* 和 *nad6*), 余下的 4 个基因 (*nad1*、*nad4*、*nad4L* 和 *nad5*) 则在轻链上编码 (表 3, 图 2)。南极磷虾 (Prydz Bay) 线粒体基因组 13 个蛋白质编码基因转录的方向和相对位置, 与泛甲壳动物线粒体基因组中这些基因的方向和位置完全一致 (图 2)。

南极磷虾 (Prydz Bay) 线粒体基因组的 13 个蛋白质编码基因中, 出现两个阅读框的重叠 (两对基因: *atp6-atp8* 和 *nad4-nad4L* 均有 7 个碱基的重叠), 这与甲壳动物线粒体基因组的典型特征相同 (表 3, Shen *et al.*, 2007)。在南极磷虾 (Prydz Bay) 线粒体基因组中, 5 个蛋白质编码基因 (*atp6*、*cob*、*cox3*、*nad4* 和 *nad4L*) 以 ATG 作为起始密码子, 而 *cox2* 和 *nad1* 基因的起始密码子为 ATA, *nad2*、*nad3* 和 *nad6* 基因以 ATT 起始, *atp8*、*cox1* 和 *nad5* 基因起始密码子分别为 ATC、ACG 和 ATG (表 3), 这和南极磷虾 (Weddell Sea) 完全一致 (Machida *et al.*, 2004)。

南极磷虾 (Prydz Bay) 线粒体基因组 13 个蛋白质编码基因中, 除了 *cox2*、*cox3* 和 *nad2* 为不完全终止密码子 (TA-或 T-) 外, 其余的 10 个蛋白质编码基因均以 TAA 或者 TAG 终止 (TAA : *atp6*、*atp8*、*cob*、*cox1*、*nad3-4*、*nad4L*、*nad5* 和 *nad6*; TAG : *nad1*)。南极磷虾 (Weddell Sea) 和南极磷虾 (Prydz Bay) 相比, 还出现 6 处改变, 包括 *atp6*、*atp8*、*nad2*、*nad3*、*nad4L*、*nad5* 和 *nad6* (Machida *et al.*, 2004)。这些不同, 部分是由于

表 3 南极磷虾(软甲纲:磷虾目)线粒体基因组的基因分布特征
Tab.3 Mitochondrial gene profile of *Euphausia superba* (Malacostraca: Euphausiacea)

基因名称	位置	编码链	长度(bp)	密码子		基因间区长度*
				起始	终止	
nCR	1—316	+	316			
<i>tRNA</i> ^{Asn}	317—381	+	65			0
<i>tRNA</i> ^{Gln}	379—447	-	69			-3
<i>tRNA</i> ^{Met}	447—514	+	68			-1
<i>nad2</i>	515—1514	+	1000	ATT	T-	0
<i>tRNA</i> ^{Cys}	1515—1578	-	64			0
<i>tRNA</i> ^{Tyr}	1590—1656	-	67			11
<i>tRNA</i> ^{Trp}	1672—1741	+	70			15
<i>cox1</i>	1746—3284	+	1539	ACG	TAA	4
<i>tRNA</i> ^{Leu(CUN)}	3298—3363	+	66			13
<i>cox2</i>	3364—4051	+	688	ATA	T-	0
<i>tRNA</i> ^{Lys}	4052—4120	+	69			0
<i>tRNA</i> ^{Asp}	4121—4188	+	68			0
<i>atp8</i>	4189—4347	+	159	ATC	TAA	0
<i>atp6</i>	4341—5015	+	675	ATG	TAA	-7
<i>cox3</i>	5015—5807	+	793	ATG	T-	-1
<i>tRNA</i> ^{Gly}	5808—5874	+	67			0
<i>nad3</i>	5875—6228	+	354	ATT	TAA	0
<i>tRNA</i> ^{Ala}	6228—6293	+	66			-1
<i>tRNA</i> ^{Arg}	6295—6361	+	67			1
<i>tRNA</i> ^{Asn}	6362—6426	+	65			0
<i>tRNA</i> ^{Ser(AGN)}	6426—6493	+	68			-1
<i>tRNA</i> ^{Glu}	6496—6564	+	69			2
<i>tRNA</i> ^{Phe}	6664—6731	-	68			99
<i>nad5</i>	6731—8461	-	1731	ATG	TAA	-1
<i>tRNA</i> ^{His}	8462—8527	-	66			0
<i>nad4</i>	8528—9865	-	1338	ATG	TAA	0
<i>nad4L</i>	9861—10158	-	298	ATG	TAA	-5
<i>tRNA</i> ^{Thr}	10161—10226	+	66			2
<i>tRNA</i> ^{Pro}	10228—10294	-	67			1
<i>nad6</i>	10298—10819	+	522	ATT	TAA	3
<i>cob</i>	10819—11955	+	1137	ATG	TAA	-1
<i>tRNA</i> ^{Ser(UCN)}	11976—12046	+	71			20
<i>nad1</i>	12064—13002	-	939	ATA	TAG	17
<i>tRNA</i> ^{Leu(UUR)}	13019—13084	-	66			16
<i>lrRNA</i>	13085—14410	-	1326			0
<i>tRNA</i> ^{Val}	14411—14482	-	72			0
<i>srRNA</i>	14483—15290	-	808			0
<i>tRNA</i> ^{Ile}	15291—15357	+	67			0
nCR	15358—15498	+	141			0

* 数字表示基因间区的长度, 负值表示相邻基因之间存在重叠。“-”表示通过转录后的多腺苷酸化来形成完整的终止密码子

表 4 南极磷虾线粒体基因组的基因组特征
Tab.4 Mitochondrial genomic characteristics of *Euphausia superba*

GenBank 接收号	主编码链		13 个蛋白质编码基因				<i>lrRNA</i>		<i>srRNA</i>		<i>tRNA</i>			
	长度 (bp)	A+T 含量 (%)	氨基酸数量 [†]	A+T 含量(%)			长度 (bp)	A+T 含量 (%)	长度 (bp)	A+T 含量 (%)	长度 (bp)	A+T 含量 (%)		
				所有位点	第一位	第二位							第三位	
<i>E. superba</i> (Prydz Bay)	EU583500	15,498	68.1	3711	66.2	58.8	62.2	77.7	1326	75.7	808	75	1551	68.5
<i>E. superba</i> (Weddell Sea)	AB084378	14,606	67.7	3714	66.3	58.8	62.3	77.9	1326	75.8	618*	74.6	1234	67.7

注：† 不包括终止密码子；* 序列不完整



图 2 南极磷虾(软甲纲:磷虾目)线粒体基因组的基因分布

Fig.2 Gene map of mitochondrial genomes of *Euphausia superba* (Malacostraca: Euphausiacea)

注:蛋白质编码基因顺时针方向转录,除了 *nad1*、*nad4*、*nad4L* 和 *nad5* 基因(以下划线表示); 2 个核糖体 RNA 在轻链上编码(以下划线表示); 转运 RNA 以单个字母的氨基酸密码子表示,编码在重链和轻链上分别以环外和环内表示。内部的圆环显示的为 AT 含量。南极磷虾(Prydz Bay)线粒体基因组序列提交至 GenBank 的接收号为 EU583500

南极磷虾(Weddell Sea)在基因预测时产生的差异。不成熟的终止密码子在后生动物线粒体基因组中很常见,而且 Ojala 等(1981)的研究表明,成熟的终止密码子 TAA 可以通过转录后的多腺苷酸化产生。

2.3 碱基组成和密码子使用偏好

碱基组成的偏好性可以通过偏移度(skewness)来描述,这个指标能够衡量 A 对 T 和 G 对 C 的相对数量,分别以 $(A\%-T\%)/(A\%+T\%)$ 和 $(G\%-C\%)/(G\%+C\%)$ 来计算(Perna et al, 1995)。南极磷虾(Prydz Bay)线粒

体基因组主编码链的碱基组成为 33.8% A、18.3% C、13.6% G 和 34.3% T, 呈现出较弱的 T 偏好(AT skew = -0.007)和较强的 C 偏好(GC skew = -0.147)(表 5)。在南极磷虾(Prydz Bay)的蛋白质编码基因中,除了 *atp8* 基因呈现较弱的 A 偏好(AT skew = 0.042)之外,其余基因均具有不同程度的 T 偏好(AT skew 介于 -0.040 和 -0.257 之间)(表 5)。相比而言,GC 偏好的情况要复杂一些: *nad1*、*nad4*、*nad4L* 和 *nad5* 偏好于 G(GCskew=0.079—0.237), 而其他的蛋白质编码基因却偏好于 C(GCskew= -0.063 — -0.620)。南极磷虾(Prydz Bay)线粒体基因组的 *srRNA* 基因具有微弱的 A 偏好(AT skew = 0.064)和微弱的 G 偏好(GC skew=0.072); 同时 *lrRNA* 基因具有微弱的 A 偏好(AT skew = 0.025)和较强的 G 偏好(GC skew=0.202)(表 5)。

南极磷虾(Prydz Bay)线粒体基因组的密码子使用情况见表 6。在南极磷虾(Prydz Bay)线粒体基因组 13 个蛋白质编码基因中,共包含 3,711 个密码子(终止子除外)。南极磷虾(Prydz Bay)线粒体基因组中,出现频率最高的氨基酸是 Leu(15.44%), 然后依次是: Ser(9.57%)、Phe(7.90%)、Ile(7.60%)和 Val(7.33%)。后生动物线粒体基因组存在的一个普遍现象:对于代表核苷酸 A 和 T 的密码子具有一定的偏好,这主要来源于碱基的偏好。南极磷虾(Prydz Bay)线粒体基因组总的 AT 含量为 66.2%, 然而在密码子的第三位 AT 含量上升至 77.7%(表 4)。

2.4 转运 RNA 和核糖体 RNA

南极磷虾线粒体基因组共编码 23 个转运 RNA, 全长 1,551 bp, AT 含量为 68.5%(表 4); 单个基因长度为 64—72 bp(表 3), 其与后生动物线粒体基因组编码的 22 个转运 RNA 相比,则存在 1 个 *trnN* 基因重复。南极磷虾线粒体基因组编码转运 RNA 基因的长度和反密码子使用与其他已测定的大多数软甲纲动

表 5 南极磷虾线粒体基因组的碱基组成与偏好
Tab.5 Nucleotide composition and skew of the *Euphausia superba* mitochondrial genome

基因名称	碱基组成				A+T 含量(%)	AT 偏好	GC 偏好
	A	C	G	T			
<i>atp6</i>	0.298	0.211	0.135	0.356	65.4	-0.089	-0.220
<i>atp8</i>	0.397	0.192	0.045	0.365	76.2	0.042	-0.620
<i>cob</i>	0.267	0.190	0.153	0.391	65.8	-0.188	-0.108
<i>cox1</i>	0.265	0.203	0.179	0.353	61.8	-0.142	-0.063
<i>cox2</i>	0.308	0.195	0.163	0.334	64.2	-0.040	-0.089
<i>cox3</i>	0.259	0.198	0.174	0.369	62.8	-0.175	-0.065
<i>nad1</i>	0.280	0.115	0.183	0.422	70.2	-0.202	0.228
<i>nad2</i>	0.276	0.223	0.152	0.349	62.5	-0.117	-0.189
<i>nad3</i>	0.279	0.214	0.142	0.365	64.4	-0.134	-0.202
<i>nad4</i>	0.275	0.121	0.196	0.407	68.2	-0.194	0.237
<i>nad4L</i>	0.249	0.152	0.178	0.421	67.0	-0.257	0.079
<i>nad5</i>	0.287	0.129	0.176	0.408	69.5	-0.174	0.154
<i>nad6</i>	0.277	0.168	0.106	0.449	72.6	-0.237	-0.226
<i>srRNA</i>	0.399	0.116	0.134	0.351	75.0	0.064	0.072
<i>lrRNA</i>	0.388	0.097	0.146	0.369	75.7	0.025	0.202
total	0.338	0.183	0.136	0.343	68.1	-0.007	-0.147

注: 根据 Perna 等(1995)的公式: $AT\ skew = ((A\% - T\%) / (A\% + T\%))$ 和 $GC\ skew = ((G\% - C\%) / (G\% + C\%))$ 计算 AT 偏好和 GC 偏好

表 6 南极磷虾(Prydz Bay)线粒体基因组 13 个蛋白质编码基因的密码子使用情况
Tab.6 Codon usage in 13 protein-coding genes of *Euphausia superba* (Prydz Bay)

Phe	UUU	220	Ser	UCU	111	Tyr	UAU	106	Cys	UGU	34	
	UUC	73		UCC	18		UAC	43		UGC	8	
Leu	UUA	281	UUG	UCA	90	End	UAA	9	Trp	UGA	77	
	UUG	45		UCG	9		UAG	1		UGG	24	
Leu	CUU	101	Pro	CCU	62	His	CAU	42	Arg	CGU	24	
	CUC	32		CCC	13		CAC	43		CGC	2	
	CUA	97		CCA	55		Gln	CAA		46	CGA	29
	CUG	17		CCG	14		CAG	27		CGG	7	
Ile	AUU	217	Thr	ACU	89	Asn	AAU	87	Ser	AGU	37	
	AUC	65		ACC	23		AAC	40		AGC	17	
Met	AUA	170	AUG	ACA	85	Lys	AAA	64	AGA	AGA	54	
	AUG	46		ACG	11		AAG	20		AGG	19	
Val	GUU	117	Ala	GCU	107	Asp	GAU	49	Gly	GGU	98	
	GUC	18		GCC	29		GAC	33		GGC	20	
	GUA	105		GCA	98		Glu	GAA		46	GGA	87
	GUG	32		GCG	10		GAG	28		GGG	40	

物一致。南极磷虾线粒体基因组 tRNA 基因的位置和泛甲壳动物线粒体基因组的原始排列相比存在一定程度的易位(图 2)。

DOGMA 和 BLAST 的分析结果显示, 南极磷虾线粒体基因组的 *lrRNA* 基因位于 *tRNA^{Leu}(^{CUN})* 和

tRNA^{Val} 之间, 而 *srRNA* 位于 *tRNA^{Val}* 和假定的控制区之间, 并且, 两个核糖体 RNA 都在负链上编码。南极磷虾核糖体 RNA 在线粒体基因组上的位置和方向与泛甲壳动物的典型特征一致。南极磷虾 (Prydz Bay) *lrRNA* 和 *srRNA* 的长度分别为 1 326 bp 和 808 bp,

AT的含量分别为 75.7%和 75.0%，这些都和其他软甲纲的物种类似(表 4)。

2.5 南极磷虾线粒体基因组单基因的相似性分析

表 7 显示了来自普里兹湾和威德尔海南极磷虾线粒体基因组主编码基因之间的差异位点和相似度。通过比较发现, *atp8* 基因全长 156 bp, 差异位点为零; *cox1* 基因全长 1 536 bp, 并存在 3 个位点的差异。在所有的主编码基因中, 变异最大的是 *nad2* 基因, 差异位点高达 61 个(基因全长 990 bp), 这使得南极磷虾这两个类群 *nad2* 基因之间的相似度仅为 93.838%; 其次是 *nad5* 基因, 差异位点有 12 个(基因全长 1 728 bp)。另外, 两个核糖体 RNA(*srRNA* 和 *lrRNA*)的差异位点分别为 2 个和 1 个, 相似度分别高达 99.676%和 99.925%。

表 7 南极磷虾(Prydz Bay)和南极磷虾(Weddell Sea)线粒体基因组单基因之间的差异位点和相似度分析

Tab.7 Mitochondrial gene variant sites and identities between *Euphausia superba* (Prydz Bay) and *E. superba* (Weddell Sea)

基因名称	长度(bp)	差异位点*	相似度(%)
<i>atp6</i>	672	2	99.702
<i>atp8</i>	156	0	100.000
<i>cob</i>	1134	3	99.735
<i>cox1</i>	1536	3	99.805
<i>cox2</i>	687	4	99.418
<i>cox3</i>	792	3	99.621
<i>nad1</i>	936	5	99.466
<i>nad2</i>	990	61	93.838
<i>nad3</i>	351	1	99.715
<i>nad4</i>	1335	8	99.401
<i>nad4L</i>	297	2	99.327
<i>nad5</i>	1728	12	99.306
<i>nad6</i>	519	4	99.229
<i>srRNA</i>	618	2	99.676
<i>lrRNA</i>	1326	1	99.925

* 包括突变、插入和缺失

3 讨论

本研究中测定了来自南极普里兹湾(Prydz Bay)南极磷虾的线粒体基因组序列。南极磷虾(Prydz Bay)线粒体基因组全长为 15 498 bp 以上, 与后生动物线粒体基因组典型的基因组成相比, 存在 1 个 *trnN* 基因的重复, 这和大多数的甲壳动物不同(Hickerson *et al.*, 2000; Wilson *et al.*, 2000; Yamauchi *et al.*, 2002; Yamauchi *et al.*, 2003; Miller *et al.*, 2004; Yamauchi *et al.*, 2004; Miller *et al.*, 2005; Place *et al.*, 2005; Segawa *et al.*,

2005; Sun *et al.*, 2005; Kilpert *et al.*, 2006; Miller *et al.*, 2006; Ivey *et al.*, 2007; Shen *et al.*, 2007)。

线粒体基因组基因排列的比较研究, 在解决古老的系统发生关系方面正脱颖而出, 成为一种全新的、有效的研究手段(Boore *et al.*, 1995; Boore *et al.*, 1998; Shen *et al.*, 2007)。Boore 等(1995)将线粒体基因排列应用于节肢动物的系统发生研究, 发现节肢动物具有一些共享的基因排列, 为节肢动物单系起源提供强有力的证据。Boore 和 Brown(1998)通过比较 144 种节肢动物和另外 9 种原口动物线粒体基因组中 *trnL₂* 基因的相对位置, 发现共享的 *trnL₂* 易位支持昆虫与甲壳动物近缘。甲壳动物和六足动物线粒体基因组原始的基因排列完全一致, 从而支持二者构成单系的泛甲壳动物(Boore *et al.*, 1998)。然而, 在 GenBank 公布的 36 个甲壳动物线粒体基因组中, 只有 13 个物种线粒体基因组的基因排列与泛甲壳动物的原始排列一致。软甲纲线粒体基因组的基因排列呈现出很高的多样性, 其中腹胚亚目的部分类群发生不同规模的基因重排。Shen 等(2007)通过线粒体基因组的基因排列分析, 发现 5 个短尾类物种的线粒体基因组和泛甲壳动物的原始排列相比, 均存在 *trnH* 基因易位, 因此推断这种共享的基因易位为短尾下目的单系起源提供证据。南极磷虾(Prydz Bay)线粒体基因组的基因排列与泛甲壳动物线粒体基因组的原始排列相比, 存在 4 个 tRNA 基因易位和 1 个 *trnN* 基因重复。基因排列的比较对于分子系统学研究是非常重要的信息来源(Boore *et al.*, 1998), 南极磷虾(Prydz Bay)线粒体基因组的测定为从基因组水平上探讨磷虾目在软甲纲中的分类地位以及与其他类群的系统发生关系提供了条件。

磷虾类作为重要的海洋浮游动物, 在海洋生态系统中发挥着举足轻重的作用(孙松, 2002)。磷虾类分子生态学和群体遗传学的研究必将成为未来海洋浮游动物生态学研究的一个重要方面。本研究的主要目标就是揭示南极磷虾线粒体基因组中主编码基因的变异情况, 从而为南极磷虾的分子生态学和群体遗传学研究奠定基础。

通过比较来自 Prydz Bay 和 Weddell Sea 的南极磷虾线粒体基因组, 发现 *atp8* 基因的差异位点为零, 这和通常认为 *atp8* 基因存在较高的进化速率并不相同(Shen *et al.*, 2007)。*Cox1* 基因通常被用于 barcoding

分析和海洋生物类群的群体遗传学研究(Hebert *et al*, 2004; Schindel *et al*, 2005; Seifert *et al*, 2007)。本研究发现两个南极磷虾的 *cox1* 基因中仅存在 3 个位点的差异, 这在群体遗传学研究中往往是不够的。另外, 在南极磷虾所有的主编码基因中, 变异最大的是 *nad2* 基因, 差异位点高达 61 个(基因全长 990 bp), 这使得南极磷虾这两个类群 *nad2* 基因之间的相似度仅为 93.838%; 其次是 *nad5* 基因, 差异位点达到 12 个(基因全长 1 728 bp); 而 *cox1* 基因的变异位点仅有 3 个。因此, 作者认为 *nad2* 基因和 *nad5* 基因可以作为 *cox1* 基因辅助的分子标记, 用于分析南极磷虾不同地理类群之间的遗传多样性, 为合理利用南极磷虾生物资源提供更多保障。

海洋无脊椎动物线粒体基因组的研究在国内才刚刚起步, 目前仅见零星报道(Sun *et al*, 2005; Shen *et al*, 2007)。相比之下, 脊椎动物的相关研究尤其是两栖类主要类群的线粒体基因组学研究开展的较为深入(Zhang *et al*, 2003; Liu *et al*, 2005; Zhang *et al*, 2005; Zhang *et al*, 2006; Su *et al*, 2007; Zhang *et al*, 2008)。线粒体基因组无疑蕴涵了物种进化历程中重要的遗传信息, 而海洋浮游动物作为海洋生态系统中最为关键的类群, 其线粒体基因组学的研究在国内依然是空白。随着我国对海洋生物多样性和海洋生态系统研究的不断深入, 线粒体基因作为重要的分子标记, 必将受到越来越多的关注。加强我国海洋浮游动物线粒体基因组学方面的研究, 尤其是加强基因组信息空白门类及生态意义重大类群线粒体基因组的研究, 符合科学研究的发展趋势。

参 考 文 献

- 孙 松, 严小军, 2001. 南极大磷虾的生物活性物质及其用途研究进展. 极地研究, 13:213—216
- 孙 松, 2002. 南极磷虾. 世界科技研究与发展, 24:57—60
- Boore J L, Lavrov D V, Brown W M, 1998. Gene translocation links insects and crustaceans. *Nature*, 392:667—668
- Ewing B, Green P, 1998a. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8:186—194
- Ewing B, Hillier L, Wendl M C *et al*, 1998b. Base-calling of automated sequencer traces using phred I. Accuracy assessment *Genome Res*, 8:175—185
- Gordon D, Abajian C, Green P, 1998. Consed: a graphical tool for sequence finishing. *Genome Res*, 8:195—202
- Hebert P D, Stoeckle M Y, Zemplak T S *et al*, 2004. Identification of birds through DNA barcodes. *PLoS Biol*, 2:e312
- Hickerson M J, Cunningham C W, 2000. Dramatic mitochondrial gene rearrangements in the hermit crab *Pagurus longicarpus* (Crustacea, anomura). *Mol Biol Evol*, 17:639—644
- Ivey J L, Santos S R, 2007. The complete mitochondrial genome of the Hawaiian anchialine shrimp *Halocaridina rubra* Holthuis, 1963 (Crustacea: Decapoda: Atyidae). *Gene*, 394:35—44
- Kilpert F, Podsiadlowski L, 2006. The complete mitochondrial genome of the common sea slater, *Ligia oceanica* (Crustacea, Isopoda) bears a novel gene order and unusual control region features. *BMC Genomics*, 7:241
- Liu Z Q, Wang Y Q, Su B, 2005. The mitochondrial genome organization of the rice frog, *Fejervarya limnocharis* (Amphibia: Anura): a new gene order in the vertebrate mtDNA. *Gene*, 346:145—151
- Lohse M, Drechsel O, Bock R, 2007. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet*, 52:267—274
- Lowe T M, Eddy S R, 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*, 25:955—964
- Machida R J, Miya M U, Yamauchi M M *et al*, 2004. Organization of the mitochondrial genome of Antarctic krill *Euphausia superba* (Crustacea: Malacostraca). *Mar Biotechnol*, 6:238—250
- Miller A D, Austin C M, 2006. The complete mitochondrial genome of the mantid shrimp *Harpisquilla harpax*, and a phylogenetic investigation of the Decapoda using mitochondrial sequences. *Mol Phylogenet Evol*, 38:565—574
- Miller A D, Murphy N P, Burrige C P *et al*, 2005. Complete mitochondrial DNA sequences of the decapod crustaceans *Pseudocarcinus gigas* (Menippidae) and *Macrobrachium rosenbergii* (Palaemonidae). *Mar Biotechnol*, 7:339—349
- Miller A D, Nguyen T T, Burrige C P *et al*, 2004. Complete mitochondrial DNA sequence of the Australian freshwater crayfish, *Cherax destructor* (Crustacea: Decapoda: Parastacidae): a novel gene order revealed. *Gene*, 331:65—72
- Ojala D, Montoya J, Attardi G, 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature*, 290:470—474
- Perna N T, Kocher T D, 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J Mol Evol*, 41:353—358
- Place A R, Feng X, Steven C R *et al*, 2005. Genetic markers in blue crabs (*Callinectes sapidus*) II. Complete mitochondrial genome sequence and characterization of genetic variation. *J Exp Mar Biol Ecol*, 319:15—27
- Rozas J, Sanchez-DelBarrio J C, Messeguer X *et al*, 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, 19:2496—2497
- Schindel D E, Miller S E, 2005. DNA barcoding a useful tool for taxonomists. *Nature*, 435:17
- Segawa R D, Aotsuka T, 2005. The mitochondrial genome of the Japanese freshwater crab, *Geothelphusa dehaani* (Crustacea: Brachyura): evidence for its evolution via gene duplication. *Gene*, 355:28—39
- Seifert K A, Samson R A, Dewaard J R *et al*, 2007. Prospects for fungus identification using CO1 DNA barcodes, with *Penicillium* as a test case. *Proc Natl Acad Sci USA*, 104:3901—3906
- Shen X, Ren J F, Cui Z X *et al*, 2007. The complete mitochondrial genomes of two common shrimps (*Litopenaeus vannamei* and *Fenneropenaeus chinensis*) and their phylogenomic considerations. *Gene*, 403:98—109
- Su X, Wu X B, Yan P *et al*, 2007. Rearrangement of a mitochon-

- drial tRNA gene of the concave-eared torrent frog, *Amolops tormotus*. *Gene*, 394:25—34
- Sun H, Zhou K, Song D, 2005. Mitochondrial genome of the Chinese mitten crab *Eriocheir japonica sinensis* (Brachyura: Thoracotremata: Grapsoidea) reveals a novel gene order and two target regions of gene rearrangements. *Gene*, 349:207—217
- Wilson K, Cahill V, Ballment E *et al*, 2000. The complete sequence of the mitochondrial genome of the crustacean *Penaeus monodon*: are malacostracan crustaceans more closely related to insects than to branchiopods? *Mol Biol Evol*, 17:863—874
- Wyman S K, Jansen R K, Boore J L, 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, 20:3252—3255
- Yamauchi M M, Miya M U, Nishida M, 2002. Complete mitochondrial DNA sequence of the Japanese spiny lobster, *Panulirus japonicus* (Crustacea: Decapoda). *Gene*, 295:89—96
- Yamauchi M M, Miya M U, Nishida M, 2003. Complete mitochondrial DNA sequence of the swimming crab, *Portunus trituberculatus* (Crustacea: Decapoda: Brachyura). *Gene*, 311:129—135
- Yamauchi M M, Miya M U, Machida R J *et al*, 2004. PCR-based approach for sequencing mitochondrial genomes of decapod crustaceans, with a practical example from kuruma prawn (*Marsupenaeus japonicus*). *Mar Biotechnol*, 6:419—429
- Zhang H, Yan J, Zhang G *et al*, 2008. Phylogeography and demographic history of Chinese black-spotted frog populations (*Pelophylax nigromaculata*): evidence for independent refugia expansion and secondary contact. *BMC Evol Biol*, 8:21
- Zhang P, Chen Y Q, Liu Y F *et al*, 2003. The complete mitochondrial genome of the Chinese giant salamander, *Andrias davidianus* (Amphibia: Caudata). *Gene*, 311:93—98
- Zhang P, Chen Y Q, Zhou H *et al*, 2006. Phylogeny, evolution, and biogeography of *Asiatic Salamanders* (Hynobiidae). *Proc Natl Acad Sci USA*, 103:7360—7365
- Zhang P, Zhou H, Liang D *et al*, 2005. The complete mitochondrial genome of a tree frog, *Polypedates megacephalus* (Amphibia: Anura: Rhacophoridae), and a novel gene organization in living amphibians. *Gene*, 346:133—143
- Zhou M, Dorland R D, 2004. Aggregation and vertical migration behavior of *Euphausia superba*. *Deep sea research*, 51:2119—2137

MITOCHONDRIAL GENOMIC CHARACTERISTICS OF *EUPHAUSIA SUPERBA* AND ITS APPLICATION AS MOLECULAR MARKERS

SHEN Xin^{1,3}, SUN Song^{1,3}, WANG Hai-Qing^{1,3}, WANG Min-Xiao^{1,3}, REN Jian-Feng^{1,3}, ZHANG Guang-Tao¹, LIU Bin^{1,2,3}

(1. Institute of Oceanology, Chinese Academy of Sciences, Qingdao, 266071;
2. Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, 100029;
3. Graduate University, Chinese Academy of Sciences, Beijing, 100039)

Abstract To reveal mitochondrial genomic characteristics of *Euphausia superba* (Crustacea: Malacostraca: Euphausiacea) and potential molecular markers, the mitochondrial genomes of *E. superba* sampled from different marine areas of the Antarctica were compared. The mitochondrial genomic DNA of *E. superba* sampled from Prydz Bay, was obtained with Long PCR technique, and then was sequenced using primer-walking and shotgun strategies. Sequence analysis and gene annotation were carried out with bioinformatic software DOGMA, tRNAscan-SE 1.21, and DnaSP 4.10.7 etc. Having its partial non-coding region completely determined, more than 15498 bp long mitochondrial genome of the Antarctic *E. superba* was revealed, containing a set of 13 protein-coding genes, 2 ribosomal RNA genes, and 23 transfer RNA genes. An extra *trnN* gene was also found by comparing it with the standard set of metazoan mitochondrial genomes. Translocation of four tRNAs (*trnL*₁, *trnL*₂, *trnW* and *trnI*) and one duplication of *trnN* were found in the mitochondrial genome. After compared two populations of *E. superba* in Prydz Bay and the Weddell Sea in mitochondrial genome, it was found that *nad2* gene has maximal variation in 61 different sites, followed by *nad5* gene for 12, while *cox1* gene has only 3. Thereby, *nad2* gene and *nad5* gene are accessory molecular markers to *cox1* gene, and be used to study the inherited diversity among different *E. superba* populations for better exploitation and management of this bio-resource.

Abbreviations: *atp6*, and 8, ATPase subunits 6 and 8; bp, base pair (s); *cox1*—3, cytochrome c oxidase subunits I—III; PCGs, protein coding genes; nCR, non coding region; *cob*, cytochrome *b*; mtDNA, mitochondrial DNA; *nad1*—6, and 4L, NADH dehydrogenase subunits 1—6 and 4L; *srRNA*, and *lrRNA*, small and large subunits ribosomal RNA; *tRNA*, transfer RNA; *L*₁, *tRNA*^{Leu(CUN)}, *L*₂, *tRNA*^{Leu(UUR)}, *S*₁, *tRNA*^{Ser(AGN)}, *S*₂, *tRNA*^{Ser(UCN)}.

Key words *Euphausia superba*, Mitochondrial genome, Marine ecology, Zooplankton, Molecular marker