

# 河姆渡古稻 DNA 提取及其序列分析

樊龙江<sup>①</sup>, 桂毅杰<sup>①</sup>, 郑云飞<sup>②</sup>, 王煜<sup>①</sup>, 蔡大广<sup>③</sup>, 游修龄<sup>④</sup>

① 浙江大学农学系, 杭州 310058;

② 浙江省文物考古研究所, 杭州 310014;

③ Department of Molekulare Phytopathologie, Christian-Albrechts-Universität zu Kiel, D-24118 Kiel, Germany;

④ 浙江大学人文学院, 杭州 310028

E-mail: fanlj@zju.edu.cn

2011-02-22 收稿, 2011-07-11 接受

国家重点基础研究发展计划(2011CB109306)、国家高技术研究发展计划(2006AA10A102)、科技部专项(2007DKA20Z90)和德国 DAAD 学者资助项目

**摘要** 水稻(*Oryza sativa*)最初在长江中下游地区被驯化, 在中国境内已发现大量稻作遗址。目前稻作史研究均是根据形态如小穗轴等进行驯化进程分析, 至今尚无利用序列数据进行分析的报道。本研究对来自 3 个考古遗址(浙江田螺山遗址、江西新干战国粮仓和浙江湖州毗山遗址唐代文化层, 年代分别为 7000, 2400 和 1200 BP)稻作遗存进行了 DNA 提取, PCR 克隆测定了 4 个基因组片段。序列分析表明, 每个遗址均发现 2 个以上的水稻基因型, 表明当时的种质尚处于混杂状态。与来自亚洲栽培稻以及野生稻的比较分析, 发现至少一个古稻基因型在栽培稻中可能已丢失, 同时本研究稻作遗存可能为粳稻类型, 或当时可能尚处于籼粳混种或籼粳分化不明显状态, 但粳稻类型占优势。本研究结果表明水稻在驯化过程中发生了显著的基因型分化。在田螺山遗址稻作遗存中还发现与无患子目、蕨藜目和十字花目植物高度同源的序列, 这与遗址发现大量楝树遗存以及南方常见田间杂草芥类等植物遗存相吻合。

## 关键词

*Oryza sativa*  
古 DNA  
测序  
田螺山遗址  
水稻驯化

水稻(*Oryza sativa*)是亚洲最重要的粮食作物之一。栽培稻分为籼稻(*indica*)和粳稻(*japonica*) 2 个亚种。亚洲野生稻最早是在中国长江流域被驯化的<sup>[1]</sup>。一些分子群体遗传学研究表明, 水稻至少经历 2 次独立驯化起源过程, 粳稻驯化起源于中国, 而籼稻的驯化发生在喜马拉雅山以南的印度等地<sup>[2]</sup>。在中国境内发现了大量稻作遗存, 如浙江余姚河姆渡、湖南澧县和河南贾湖等地<sup>[3]</sup>。稻作史研究一个重要的事件是 20 世纪 70 年代在河姆渡新石器文化遗址中发现大量古稻谷<sup>[4]</sup>, 该古稻距今约 7000 年, 根据出土的稻壳小穗轴特征判断, 已发现大量不落粒的驯化稻谷<sup>[5,6]</sup>。古稻是研究水稻从野生稻向现代栽培稻驯化的重要中间节点, 它可以提供水稻进化进程的关键证据。

DNA 提取和测序技术的快速发展为开展分子考古学研究提供了可能。目前已有部分来自考古挖掘获得的人类和动植物遗存全基因组或基因组片段被测序, 为人类和动植物进化研究提供了重要证据。从植物遗存中进行序列测定的研究已在多个作物展开, 其中在玉米方面<sup>[7-11]</sup>做的工作较多, 在稻作遗存方面虽然做过一些 PCR 扩增分析的努力<sup>[12]</sup>, 但尚无来自古稻的分子序列数据证据。

本研究利用长江流域遗址考古挖掘获得的史前河姆渡文化时期的古稻和其他 2 个历史时期的稻谷遗存, 进行了 DNA 提取(混合样品)和 4 个基因组位点(3 个基因和 1 个基因间区段)序列的测定, 并与野生稻和栽培稻序列进行了基因型比较。

英文版见: Fan L J, Gui Y J, Zheng Y F, et al. Ancient DNA sequences of rice from the low Yangtze reveal significant genotypic divergence. Chinese Sci Bull, 2011, 56, doi: 10.1007/s11434-011-4691-9

## 1 材料与方法

(i) 材料. 本研究使用的水稻遗存材料来自长江下游地区, 包括浙江省余姚河姆渡文化遗址之一田螺山遗址等(表 1; 图 1). 为了与现代稻进行比较, 本研究还测定了 51 个亚洲栽培稻(*O. sativa*), 包括 29 个粳稻(*japonica*) (其中 6 个为浙江省粳稻地方品种)和 22 个籼稻(*indica*)品种; 15 个亚洲野生稻(*O. rufipogon*)和非洲栽培稻(*O. glaberrima*)、李氏禾(*Leersia oryzoides*) (假稻属, 由中国科学院植物研究所葛颂提供)、菰草(*Zizania latifolia*) (菰属)等与水稻相近的物种材料(表 1).

(ii) DNA 提取、测序及序列分析. 由于单一古稻种子残片中 DNA 数量有限, 本研究利用古稻遗存混合样品用于 DNA 提取. DNA 提取方法主要应用磁珠分离和纯化 DNA 技术<sup>[14]</sup>, 略有改进. DNA 分离时, 使用非离子、阳离子清洁剂以保护古 DNA 序列, 减少脱氧核糖核酸酶、多酚等杂质的影响.

本研究选择 4 对引物来扩增古 DNA 序列, 其中 2 对为植物界保守的通用引物 *Angio\_1F/Angio\_2R* 和 *HrbcL252F/HrbcL320R*<sup>[9]</sup>, 另 2 对引物分别是水稻分子标记 SSR 引物 *RM211* (<http://www.gramene.org>)和 1 个编码基因位点. 引物序列信息见表 2.

采用两轮 PCR 扩增, 25  $\mu$ L PCR 体系包括 2.5 mmol/L  $Mg^{2+}$ ; 0.2 mmol/L dNTP; 1 U *Taq* 酶(上海生工); 正反向引物各 1 mmol/L; DNA 模板 2  $\mu$ L. 扩增反应程序为: 95 $^{\circ}$ C, 5 min; 95 $^{\circ}$ C, 30 s, 55 $^{\circ}$ C, 30 s, 72 $^{\circ}$ C, 20 s, 35 个循环; 72 $^{\circ}$ C, 10 min; 10 $^{\circ}$ C 保存. 第 1 轮 PCR 产物稀释 10 倍后取 2  $\mu$ L 作为第 2 轮 PCR 反应的模板, 进行二次扩增, 第 2 轮扩增反应体系和条件不变. 扩增产物用 2% 的琼脂糖, 80 W 凝胶电泳 40 min.

Dolphin 凝胶成像系统拍照记录. 扩增产物经回收后 TA 克隆测序(上海英骏生物技术有限公司). 多序列连配使用 ClustalW ([www.ebi.ac.uk/clustalw/](http://www.ebi.ac.uk/clustalw/)). 本研究产生的来自古稻及其他序列均递交国际公共核酸序列数据库(GenBank, 登录号 JN169832~JN169947).

## 2 结果与讨论

### 2.1 DNA 提取、扩增及其验证实验

古稻遗存样品(稻壳等混杂材料)经简单无菌水清洗后放入无菌水保存. DNA 提取和扩增分别在 2 个独立实验室进行. DNA 提取在未做过水稻样品的实验室进行, 提取后的 DNA 样品在另一个独立实验室进行扩增. 经两轮扩增获得清晰的目标条带(图 2).

控制污染和进行验证实验是进行古 DNA 研究的重要环节<sup>[15]</sup>. 为了验证本扩增的真实性, 上述 DNA 样品提供给浙江大学动物科学学院动物免疫实验室(余旭平教授)进行重复实验. 该实验室从未进行过植物样品分析. 该实验室利用本研究提供的实验流程, PCR 扩增产物克隆测序 11 个, 获得 9 条目标序列. 同时, 一系列证据可以排除本研究获得的 DNA 序列来自污染的可能性: (1) 本研究多次实验获得的基因型均很多(如图 3 连配结果), 而污染的样品往往一种基因型占绝大多数; (2) 本研究获得的一些古稻基因型在现代水稻材料中并未检测到, 同时也检测到大量水稻田间环境相关杂草植物序列(见 2.2 部分), 而这些田间植物本研究所有实验室均没有开展过; (3) 本研究提取的 DNA 序列片段均比较短(<500 bp), 超过 500 bp 的引物设计均无法扩增出目标产物, 这符合古 DNA 序列特征. 古代样品由于长期暴露在环境中

表 1 本研究使用的考古遗存、现代水稻及水稻近缘物种材料

材料	材料类型	数量 <sup>a)</sup>	文化年代	年代估计	出土地点/来源
<i>O. sativa</i>	颖壳	/	河姆渡文化	~7000 年前	浙江省余姚田螺山
<i>O. sativa</i>	米粒	/	战国	~2400 年前	江西省新干县
<i>O. sativa</i>	颖壳	/	唐代	~1200 年前	浙江省湖州
<i>O. sativa</i>	幼苗	51	亚洲栽培种	现代	国际水稻研究所(IRRI) 中国水稻研究所
<i>O. rufipogon</i>	幼苗	15	亚洲野生种	现代	IRRI
<i>O. sativa</i>	/	374	亚洲栽培种	现代	文献[13]
<i>Leersia oryzoides</i>	叶片	1	野生种	现代	中国科学院植物研究所
<i>Zizania latifolia</i>	叶片	1	野生种	现代	浙江大学

a) 古稻遗存为混合样品, 具体基因型数量未知



图1 用于本研究的田螺山遗址出土的古代水稻

往往导致 DNA 链断裂而不超过 500 bp<sup>[14]</sup>。另外, 为了避免因 PCR 扩增或测序过程碱基突变造成的误差, 本研究认定一个真实突变和一个新基因型要求在一个碱基位点上至少有 2 个以上相同碱基发生或相同的 2 个克隆测序结果。

## 2.2 DNA 克隆测序结果

本研究针对河姆渡文化等 3 个稻谷遗存进行了 4 个基因组位点的克隆测序, 先后获得 142 个有效克隆测序结果。这 142 条序列, 绝大多数序列(91.5%)与已知水稻或其他植物序列一致, 表明这些序列为目标序列, 即来自水稻或环境中其他植物相应序列, 其中来自水稻为主(105 个, 80.8%)(表 3)。

为了鉴别稻作特性, 本研究应用了简并性程度不同的引物设计。本研究用的 4 个引物, 其中 2 个(引物 A 和 C)为高度保守的兼并引物设计, 是根据许多已知植物序列连配结果保守区段设计的引物<sup>[9]</sup>, 因此可以扩增出许多植物同源序列(图 3 所示引物 C 扩增产物序列, 其中包括水稻和其他植物序列)。其他两个引物来自本实验室, 其中引物 Y 设计在一个基因的 2 个外显子上, 所以在水稻近缘物种中具有一定的保守性或兼并性。引物 R 来自水稻基因间, 为 SSR 标记引物, 为水稻特异引物设计。

## 2.3 来自假稻属和菰属种子可能性

假稻属(*Leersia*)和菰属(*Zizania*)均是与水稻亲缘关系最近的物种<sup>[16]</sup> (www.ncbi.nlm.nih.gov/Taxonomy/). 李氏禾(*Leersia oryzoides*)属于假稻属, 其种子与水稻种子非常相近; 菰(*Zizania latifolia*)是古代重要粮食作物之一, 由于水稻的兴起而被取代。为了排除我们使用的稻谷遗存来自假稻属和菰属种子可能性, 利用本研究 4 对引物分别对以上 2 个属物种进行了 PCR 扩增(图 4)。结果表明, 水稻特异性引物 R 未扩出条带, 另 3 个引物均扩增出条带。进一步对这 3 个引物扩增产物进行测序。测序结果表明, 引物 A 在李氏禾和菰草上的基因型分别为 T-A-A 和 T-G-A, 不同于稻属(包括非洲稻); 两物种在引物 C 位点的基因型均为

表 2 本研究使用的引物序列

引物名称	代码	基因	正向引物序列(5'→3')	反向引物序列(5'→3')	产物长度 (bp)	来源
Angio	A	18S rDNA	TGCAGTAAAAAGCTCGTAG	GCACTCTAATTTCTTCAAAG	159	[9]
HbcL	C	rbcL	TAGCGGCGGAATCTTCTACT	TATGATAGCATCGTCGTTTG	89	[9]
RM211	R	基因间	CCGATCTCATCAACCAACTG	CTTACGAGGATCTCAAAGG	161	AP005647
YANG4	Y	CDK 抑制子	AGAGCTGGAAGCGTTCTTCG	GGCAGTCATTACAGGATCAAAG	230	Os10g33310

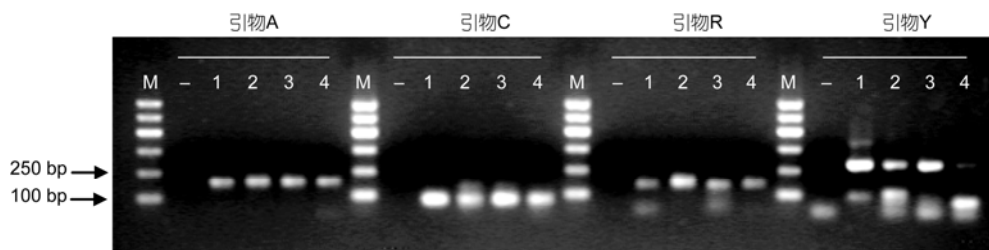


图 2 4 对引物在 4 份古稻 DNA 样品中的 PCR 凝胶电泳图

M 为 DL2000 (条带由大到小依次为 2000, 1000, 750, 500, 250 和 100 bp); “-”代表阴性对照。1~4 依次为湖州、田螺山(2 份)和新干遗存样品



表4 古稻两个基因位点克隆测序获得的序列数量分布以及与现代水稻的比较<sup>a)</sup>

引物	基因型	考古遗址				亚洲栽培稻(1)			亚洲野生稻	亚洲栽培稻(2)
		田螺山	新干	湖州	合计	籼稻	粳稻	合计		籼稻
R	A-T-A-(TC) <sub>4</sub>	12	7	9	28	9	18	27	1	17
	C-T-A-(TC) <sub>4</sub>	0	0	0	0	0	0	0	0	210
	C-C-A(TC) <sub>4</sub>	0	0	0	0	0	0	0	0	121
	A-C-T-(TC) <sub>8</sub>	0	6	2	8	0	0	0	0	0
	A-C-T-(TC) <sub>9</sub>	5	0	0	5	2	1	3	1	0
	其他	0	0	0	0	1	0	1	13	26
	合计	17	13	11	41	12	19	31	15	374
Y	T-G	12	5	1	18	1	15	16	2	0
	C-G	0	0	0	0	11	4	15	4	345
	C-C	2	1	0	3	0	0	0	0	0
	C-A	0	0	0	0	0	0	0	0	28
	合计	14	6	1	21	12	19	31	6	373

a) 亚洲栽培稻(1)和(2)数据分别来自本研究 and Huang 等人[13]

这一结果和其他研究<sup>[6]</sup>对稻谷形态观察结果一致. 每个遗址出土的稻谷外形并非完全一致, 部分稻谷间存在明显差异.

为了比较古稻基因型与现代亚洲水稻(包括栽培和野生稻)的差异, 本研究选取亚洲范围内栽培稻和野生稻对相应基因组位点进行测序, 并利用刚发表的中国水稻地方品种基因组大规模重测序数据<sup>[16]</sup>进行了基因型比较(表4). 结果表明, 古稻主要基因型(R: A-C-T-(TC)<sub>4</sub> 和 Y: T-G)均可以在现代栽培稻中找到, 说明这些基因型在栽培群体被保留下来, 同时这些基因型也是栽培粳稻的主要基因型(浙江省6个地方粳稻品种均属于这个基因型). 但同时, 古稻与现代稻基因型分布存在显著差异: (1) 古稻基因型(R: A-C-T-(TC)<sub>8</sub> 和 Y: C-C)在现代栽培稻中没有检测到, 可能在现代栽培稻遗传群体中已丢失, 有待今后更多品种特别是粳稻品种的调查加以证实. 作物在驯化过程中往往经历强烈人工选择, 遗传多态性明显下降<sup>[17]</sup>. 水稻同样经历了这样的过程, 其驯化过程中多态性下降明显, 经历了强烈的所谓选择瓶颈效应<sup>[18]</sup>. 本研究结果在水稻上直接证实了这一现象. (2) 现代栽培稻一些基因型, 特别是中国籼稻地方品种中的主要基因型(R: C-T-A-(TC)<sub>4</sub>, C-C-A(TC)<sub>4</sub> 和 Y: C-G)在古稻中均没有发现. 这可能是本研究调查古稻遗存数量的局限性造成的, 或者这些基因型可能来自新的突变或来自其他原始水稻群体. 例如, 在中

国历史上曾引进大量外来作物或作物品种, 包括水稻品种, 比较著名是宋代从越南引进的早熟稻种占城稻, 在当时江南地区大面积种植<sup>[19,20]</sup>.

籼粳分化是栽培稻驯化的一个重要进化事件. 一个有趣的问题是本研究的稻作遗存是粳稻还是籼稻. 根据本研究获得的现有两个序列片段和对籼粳稻相同基因型比率判断, 本研究的古稻序列与绝大多数粳稻一致(包括本研究调查的6个浙江省粳稻地方品种), 与绝大多数中国籼稻地方种基因型不同, 所以, 本研究所用稻作遗存可能为粳稻类型, 或至少绝大多数为粳稻类型(即当时可能处于籼粳混种或籼粳分化不明显状态).

### 2.5 来自其他植物情况

利用2个兼并引物(引物A和C), 在本研究克隆测序获得的序列中(表3), 除大多数能与水稻匹配上外(图3), 其他25条与其他植物已知序列具有显著的相似度, 而且它们大多序列间只有1~2个碱基差异, 表明它们可能来自稻作环境中的其他植物. 仔细查对这些匹配序列来源, 发现这些序列主要来自无患子目(Sapindales)、蕁藜目(Zygophyllales)和十字花目(Brassicales)植物(表5). 无患子目包括楝树、桔或枳树(Citrus)等, 这些植物在江南一带均有分布. 本文作者参与河姆渡田螺山遗址中植物遗存的鉴定, 从中可以发现不少楝树果实, 这与本研究的序列结果

表5 2个兼并引物A和C获得的25条非水稻序列情况

引物代码	基因型	克隆数	遗址来源 <sup>a)</sup>	最佳匹配物种 <sup>b)</sup>	分类
A	A96	2	T/X	<i>Citrus trifoliata</i> 等( $1 \times 10^{-74}$ )	Sapindales
	A105	6	T/X	<i>Sinapis alba</i> 等 ( $4 \times 10^{-69}$ )	Brassicales
	Y8-C	1	T	<i>Cercomonas media</i> 等( $2 \times 10^{-79}$ )	Rhizaria
	A99	1	T	<i>Pichia kluyveri</i> 等( $7 \times 10^{-62}$ )	Fungi
C	C6-C	7	T/X/H	<i>Tribulus Terrestris</i> 等( $2 \times 10^{-34}$ )	Zygophyllales
	C9-5	1	T	<i>Guarea glabra</i> 等( $2 \times 10^{-34}$ )	Sapindales
	C62	1	H	<i>Peganum harmala</i> 等( $2 \times 10^{-31}$ )	Sapindales
	C61	4	T/X/H	<i>Arabidopsis thaliana</i> 等( $2 \times 10^{-34}$ )	Brassicales
	C64	1	H	<i>Coix lacryma-jobi</i> 等( $3 \times 10^{-38}$ )	Poales

a) T, 田螺山; X, 新干; H, 湖州; b) 括号内为数据库搜索BLASTN显著性E值

相符合。与蒺藜目最匹配的植物物种包括中国南方常见的杉木(*Tribulus*)。另一类十字花目植物中应该属于稻田伴生的杂草，如南方田间常见的白芥 (*Sinapis alba*)和拟南芥(*Arabidopsis thaliana*)等。此外，一条序列应该来自南方常见的薏苡(*Coix lacryma-jobi*)，两条序列可能来自植物病虫害。

### 参考文献

- Vaughan D A, Lu B, Tomooka N. The evolving story of rice evolution. *Plant Sci*, 2008, 174: 394-408
- Londo J P, Chiang Y C, Hung K H, et al. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proc Natl Acad Sci USA*, 2006, 103: 9578-9583
- 游修龄. 中国农业通史(原始社会卷). 北京: 中国农业出版社, 2008. 175-198
- 游修龄. 对河姆渡遗址第四文化层出土稻谷和骨稻的几点看法. *文物*, 1976, (8): 20-23
- 郑云飞, 孙国平, 陈旭高. 7000年前考古遗址出土稻谷的小穗轴特征. *科学通报*, 2007, 52: 1037-1041
- Fuller D Q, Qin L, Zheng Y F, et al. The domestication process and domestication rate in rice: Spikelet Bases from the Lower Yangtze. *Science*, 2009, 323: 1607-1610
- Goloubinoff P, Pääbo S, Wilson A. Evolution of maize inferred from sequence diversity of an *Adh2* gene segment from archaeological specimens. *Proc Natl Acad Sci USA*, 1993, 90: 1997-2001
- Deakin W J, Rowley-Conwy P J, Shaw C H. Amplification and sequencing of DNA from preserved sorghum of up to 2800 years antiquity found at Qasr Ibrim. *Ancient Biomolecules*, 1998, 2: 27-41
- Rollo F, Ubaldi M, Ermini L, et al. Otzi's last meals: DNA analysis of the intestinal content of the Neolithic glacier mummy from the Apls. *Proc Natl Acad Sci USA*, 2002, 99: 12594-12599
- Jaenicke-Despres V, Buckler E S, Smith B D, et al. Early allelic selection in maize as revealed by ancient DNA. *Science*, 2003, 302: 1206-1208
- Erickson D L, Smith B D, Clarke A C, et al. An Asian origin for a 10000-year-old domesticated plant in the Americas. *Proc Natl Acad Sci USA*, 2005, 102: 18315-18320
- Nakamura I, Sato Y I. Amplification of DNA fragments isolated from a single seed of ancient rice (AD 800) by polymerase chain reaction. *Chinese J Rice Sci*, 1991, 5: 175-179
- Huang X H, Wei X H, Sang T, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet*, 2010, 42: 961-976
- Levison P R, Badger S E, Dennis J, et al. Recent developments of magnetic beads for use in nucleic acid purification. *J Chromatogr A*, 1998, 816: 107-111
- 杨东亚. 古代DNA研究中污染的控制和识别. *人类学学报*, 2003, 22: 163-173
- Ge S, Sang T, Lu B R, et al. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc Natl Acad Sci USA*, 1999, 96: 14400-14405
- Doebley J F, Gaut B S, Smith B D. The molecular genetics of crop domestication. *Cell*, 2006, 127: 1309-1321
- Zhu Q, Zheng X, Luo J, et al. Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. *Mol Biol Evol*, 2007, 24: 875-888
- 曾雄生. 宋代的双季稻. *自然科学史研究*, 2002, 21: 255-268
- 游修龄. 占城稻质疑. *农业考古*, 1983, (1): 25-31