

文章编号:0427-7104(2004)06-1061-06

# 广西瑶族三个群体线粒体遗传多态分析

高 嵩<sup>1</sup>, 毛显贇<sup>1</sup>, 文 波<sup>1</sup>, 李 辉<sup>1</sup>, 宋秀峰<sup>1</sup>,  
黄文驹<sup>2</sup>, 张有隼<sup>2</sup>, 金建中<sup>1</sup>, 金 力<sup>1</sup>.

(1. 复旦大学 生命科学学院 现代人类学研究中心, 上海 200433; 2. 广西民族学院  
民族学人类学研究所, 南宁 530006)

**摘 要:** 瑶族是苗瑶系统中的一个民族, 拥有数量众多的支系, 广泛分布于中国西南地区. 对广西瑶族的三个支系(蓝靛瑶、平地瑶、盘瑶)共 100 个个体样本的 mtDNA 高变 区序列以及 13 个 RFLP 位点进行了基因分型, 并与全国各地其他少数民族、汉族群体的数据进行比较分析. 得出结论: 瑶族在线粒体遗传结构上属于南方聚类, 并且接近地理位置邻近的周边民族群体.

**关键词:** 瑶族; 线粒体; 遗传多态; 盘瑶; 蓝靛瑶; 平地瑶

**中图分类号:** Q 986 **文献标识码:** A

人类遗传物质—DNA 分子记录了人群演化的历史, 分析在不同人群里遗传标记的种类和频率的差别, 可为研究人群起源、分化、迁徙、融合等历史问题提供遗传学上的证据. 目前人类群体遗传学研究的对象包括双系遗传的常染色体 DNA 以及单系遗传的 Y 染色体 DNA 和 mtDNA. mtDNA 由于具有突变速率快, 在群体内差异丰富, 所含信息量大; 而且严格遵从母系遗传规律, 没有重组等优点, 使得 mtDNA 多态成为研究人群遗传关系的一个重要的遗传标记. 近 20 年来, 国内外相关学者在世界各个人群 mtDNA 遗传多态研究上作了大量的工作. 研究中常用的实验手段包括基于限制性片段长度多态 (RFLP) 以及 mtDNA 序列测定. 其中序列测定大多集中于突变速率相对较高, 蕴含信息较为丰富的高变 区和 区. 为了探讨苗瑶群体演化的历史, 对广西的三个瑶族支系群体(蓝靛瑶、平地瑶、盘瑶)的共 100 个个体 mtDNA 样本进行了 RFLP 分型、高变 区序列测定, 加以分析, 以期能为进一步研究苗瑶群体的历史提供资料.

瑶族历史上依山而居, 游耕迁徙的习俗传统, 使其十分容易产生众多的相对隔离的小群体. 习惯上根据各类特征将这些支系群体加以分类命名. 如依据祖先崇拜命名“盘瑶”等, 根据服装衣饰命名“蓝靛瑶”、“白裤瑶”等, 根据居住特性分为“过山瑶”, “平地瑶”等, 根据居住地点分为“东山瑶”, “八排瑶”等. 这些大多是外人对于瑶族的称呼, 并不确切, 有不少的重叠和混淆. 根据语言上的分类, 瑶族内部在语言上主要分为三个主要群体: 属于苗瑶语族 (Hmong-Mien) 勉语支 (Mienic) 的“勉” (Mian-Jin), 占整个瑶族的 70% 以上, 具有奉行盘瓠祖先崇拜的文化特征; 属于苗瑶语族苗语支的“布努”, 这支群体占整个瑶族的 20%, 共同信奉创世女神“米洛陀”女神, 在语言和文化习俗上更接近苗族; 第三支是使用侗台语族侗水语的“拉伽”. 本次研究的蓝靛瑶、平地瑶、盘瑶, 都是属于勉语支下信奉盘瓠崇拜的群体. 蓝靛瑶使用勉语的地方方言“金门 (Kim Mun)”, 平地瑶大多已改使用汉语, 但从历史渊源和文化习俗上分类, 仍是属于勉语支下的瑶族群体.

## 1 材料和方法

### 1.1 样本采集和处理

\* 收稿日期: 2004-04-05

基金项目: 国家自然科学基金九五重大项目 (39993420)

作者简介: 高 嵩 (1979—), 男, 硕士研究生; 通讯联系人金建中教授.

在广西壮族自治区调查 3 个瑶族群体. 在知情同意的原则下,每一群体随机选取身体健康,相互之间无亲表关系,三代内未和外族通婚,并居住在当地的个体. 每个个体抽取外周血 5 $\mu$ L,ACD 抗凝,常规酚-氯仿法抽提 DNA.

表 1 样品采集地点数量及族群分布

Tab. 1 Location and sample size

群体	使用语言	采集地点	个体样本量
蓝靛瑶	金门 Kim Mun	广西田林	26
平地瑶	金门 Kim Mun	广西富川	42
盘瑶	勉 Iu Mien	广西田林	32

## 1.2 DNA 多态的基因分型

mtDNA 高变区(HVS-1)片段使用 L15974 和 H16488 引物进行 PCR 扩增. PCR 反应条件如下:94 预变性 3 min;94 变性 30 s,62 退火 40 s,72 延伸 40 s;30 个循环;循环结束 72 延伸 5 min,反应体系 25  $\mu$ L. PCR 产物纯化后用 BIG-DYE 末端荧光标记试剂盒(PE 公司)进行测序反应,反应用 0.8  $\mu$ mol/L 单向引物 2  $\mu$ L,MIX 1  $\mu$ L,总体系 6  $\mu$ L. 反应产物经酒精沉淀、干燥溶于甲酰胺变性后,用 ABI3100 序列分析仪(PE 公司)进行序列分析. 序列先从 L15974 一端正向测序,一旦遇上 polyC 片段导致无法测通,则从 H16488 一端进行反向测序,序列分析比对时再将两段序列拼接完整.

除了进行高变 I 区测序外,另外使用高分辨率 RFLP 进行线粒体基因分型. 根据经修正的剑桥标准序列(revised Cambridge reference sequence,rCRS)设计引物,扩增包含所需突变位点的特定 DNA 片段,进行限制性内切酶酶切,判断基因分型. 参照 Wallace 等<sup>[1]</sup>,Ingman 等<sup>[2]</sup>,Yao 等<sup>[3]</sup>构建的东亚人群线粒体单倍群系统树. 我们选取 10397,5176,13259,12406,10310,4715,663,9824,3391,12705,9 bp del,5417,4831 等多态位点,以上几个位点包含了绝大多数东亚-中国人群线粒体基因型. 9 bp del 位点 PCR 产物直接使用质量分数为 3%的 Nussive 凝胶电泳检测. 其余多态位点的 PCR 产物使用相应限制性内切酶酶切,酶切产物经由质量分数为 3%的琼脂糖凝胶电泳检测. 如果多态位点本身不能形成酶切位点,则通过设计引物,引入酶切位点. 引物由上海生工公司合成,限制性内切酶使用 New England Biolab 公司产品.

## 1.3 数据处理

高变区序列分析比对使用 DNASTAR 软件,分析比对时参照剑桥标准序列(rCRS)校正机器误读,另外将 L15974 和 H16488 正向反向测序结果拼接完整. 用 ARLEQUIN 软件计算单倍型频率、群体  $F_{st}$  和净遗传距离  $d_A$ . 按照 Yao 等<sup>[4]</sup>,Macaulay 等<sup>[5]</sup>,Kivisild 等<sup>[6]</sup>的方法,依据高变区序列特异位点加上 RFLP 判断样品所属单倍群. 使用 SPSS 软件进行 mtDNA 单倍群频率主成分分析,制作主成分坐标图,计算主成分系数. 在 Mega2 软件平台上,使用 ME 法计算净遗传距离  $d_A$  矩阵,构建群体分子系统树.

## 2 结 果

### 2.1 mtDNA 单倍型频率

根据高变区上的特异位点以及 RFLP 可以判断样本所属的单倍群,表 2 即是三个瑶族群体的单倍群分类列表. 依照 Yao 等构建的单倍群系统树<sup>[3]</sup>,大的单倍群下又可分为小的单倍群分支,比如 M 下可分为 D、G、M8、M7 等,D 下又可分为 D4a、D4b、D5 等分支,同一单倍型群下的亚群分支具有较近的系统演化关系,所以在表 3 中同时列出 B、D、F、M7 几个较大的单倍型群总的频率统计. 从表中可看出,蓝靛瑶、盘瑶、平地瑶有着大致相同的较高的单倍群 B,盘瑶、平地瑶中有较高的 F,蓝靛瑶、平地瑶中有较高的 M7. B、F 和 M7 属于南方群体高频率单倍群,在东亚从南至北呈频率递减趋势. 另外,三个群体中都没有或只有较少的 A、D、M8 等在北方群体频率较高的单倍群. 这些使得这三个瑶族群体在遗传结构上接近于周边的中国南方人群.

表 2 根据高变 区多态位点判断的线粒体单倍型频率组频率

Tab. 2 Haplogroup frequency based on HVS-1 sequence and RFLP SNP

				( % )			
Haplogroup	盘瑶	蓝靛瑶	平地瑶	Haplogroup	盘瑶	蓝靛瑶	平地瑶
A		3		F1a1	4	3	
B4a		6	9	F1b			9
B4a1a	4	3		F3		3	
B4a3	8	3		F 总计	27	6	27
B4b1	4	3	9	G2	4		9
B4f	4			M *			9
B5a	8	10		M7		3	
B 总计	28	25	18	M7b	4	10	9
C		10		M7b1	4	16	9
C5	12	3	9	M7 总计	8	29	18
D4	4	3		M8		3	
D5	4			N9a		6	
D 总计	8	3	0	R *	8	6	9
F	4		9	R9b	8		
F1a	19		9	R9c		3	

2.2 遗传距离矩阵

用 Arlequin 软件处理高变 区序列数据,计算群体间遗传距离  $d_A$  和  $F_{st}$ 。除了实验所得盘瑶、平地瑶、蓝靛瑶数据,另外引用了其他已发表高变 区序列数据,哈萨克、藏族、蒙古族、壮族、傣族、白族、佤族<sup>[7]</sup>,云南汉族、武汉汉族、青岛汉族、辽宁汉族和广东汉族<sup>[4]</sup>,共 15 个群体计算遗传距离矩阵。可以看出无论是  $d_A$  值或是  $F_{st}$  值,盘瑶、蓝靛瑶、平地瑶之间遗传距离较小,和南方汉族、傣族、白族、壮族等邻近群体遗传距离同样较小,而和藏族、蒙古族、哈萨克等群体之间遗传距离较大。

表 3 根据高变 区序列计算的净遗传距离(  $d_A$ ) 和  $F_{st}$

Tab. 3  $d_A$  and  $F_{st}$  based on HVS-1 sequence

	盘瑶	蓝靛瑶	平地瑶	哈萨克	藏族	蒙古族	壮族	傣族	白族	佤族	云南汉	武汉汉	青岛汉	辽宁汉	广东汉
盘瑶		0.022	0.015	0.067	0.14	0.088	0.005	-0.004	0.018	0.114	-0.005	0.032	0.031	0.021	-0.004
蓝靛瑶	0.153		0.019	0.077	0.182	0.096	0.02	0.016	0.04	0.064	0.023	0.03	0.045	0.054	0.023
平地瑶	0.136	0.184		0.057	0.112	0.056	0.024	0.02	0.019	0.084	0.011	0.025	0.026	0.024	0.005
哈萨克	0.483	0.61	0.519		0.106	0.048	0.048	0.084	0.028	0.102	0.058	0.013	0.032	0.02	0.054
藏族	0.956	1.352	0.917	0.66		0.068	0.123	0.158	0.075	0.187	0.112	0.083	0.039	0.05	0.093
蒙古族	0.656	0.822	0.572	0.325	0.369		0.07	0.091	0.035	0.147	0.057	0.045	0.02	0.021	0.057
壮族	0.042	0.144	0.202	0.367	0.887	0.57		-0.003	0.016	0.071	-0.006	0.006	0.018	0.024	0.009
傣族	-0.024	0.117	0.171	0.652	1.161	0.734	-0.026		0.024	0.099	-0.007	0.03	0.038	0.039	0.004
白族	0.122	0.291	0.164	0.189	0.467	0.252	0.121	0.179		0.086	0.014	0.007	-0.007	0.003	0.005
佤族	0.79	0.418	0.691	0.672	1.226	0.974	0.502	0.707	0.572		0.095	0.052	0.079	0.106	0.066
云南汉	-0.041	0.192	0.109	0.459	0.786	0.487	-0.048	-0.059	0.097	0.689		0.01	0.011	0.006	-0.007
武汉汉	0.238	0.25	0.236	0.099	0.548	0.377	0.042	0.233	0.045	0.346	0.083		-0.01	0.002	0.007
青岛汉	0.227	0.372	0.238	0.234	0.229	0.169	0.135	0.291	-0.054	0.529	0.086	-0.08		-0.013	0.007
辽宁汉	0.152	0.453	0.227	0.145	0.3	0.174	0.184	0.301	0.019	0.74	0.051	0.013	-0.095		0.007
广东汉	-0.026	0.183	0.046	0.414	0.639	0.471	0.072	0.027	0.033	0.458	-0.053	0.053	0.052	0.057	

注:表格对角线左下角为净遗传距离(  $d_A$  ),对角线右上角为  $F_{st}$

2.3 主成分分析

图 1 是 mtDNA 单倍群频率主成分坐标图,除了盘瑶、蓝靛瑶、平地瑶三个群体的数据外,为了加以比较另外引用了本实验室已发表的白族数据<sup>[8]</sup>,Yao 等发表的云南汉族、武汉汉族、青岛汉族、辽宁汉族、广

东汉族单倍群频率数据<sup>[4]</sup>,以及根据 Yao 等发表的藏族、蒙古族、佤族、壮族、傣族 mtDNA 高变 I 区序列<sup>[7]</sup>推断的单倍群频率,不同地区不同语系的共 15 个群体进行主成分分析. 考虑到样本量不大,较大的单倍群下各个单倍群亚群分支(如 B4b1)未必能被完全覆盖,单独作为变量可能会引入误差,故而运算时将 B、D、F、M7 单倍群下的亚群分别合并为一个变量(例如 B 总计)计算. 运算结果,第一第二主成分总共占总体方差的 76.4%. 从图 1 中可以看出这 15 个群体按照地域分布,分成了南北两个大的集合,三个瑶族群体都分布在南方集合中,彼此之间相互靠近,同时与壮族、傣族、南方汉族等周边民族群体接近. 在图 1 中,线粒体遗传距离与地理空间分布密切相关,呈正比关系,和三个瑶族群体同在广西的壮族,与瑶族距离最近,提示了历史上两者之间可能存在过为数不少的 mtDNA 基因交流,这也与历史学民族学上的记载相契合.

#### 2.4 分子系统树

图 2 是根据高变 I 区序列计算所得  $d_A$  遗传距离矩阵,用 ME 法做出的群体分子系统树. 除了瑶族群体外,还另外引用了 Yao 等发表的藏族、蒙古族、白族、傣族、壮族、佤族、哈萨克族<sup>[7]</sup>,云南汉族、武汉汉族、青岛汉族、辽宁汉族、广东汉族<sup>[4]</sup>等高变 I 区序列数据.

该系统树与单倍群频率主成分分析坐标图(图 1)的结果十分吻合,北方汉族与北方少数民族分布在一个聚类,南方汉族与南方少数民族分布在另一个聚类. 三个瑶族群体都分布在南方聚类当中,互相之间距离接近,并且同壮族、傣族等地理距离靠近的周边民族群体关系紧密.

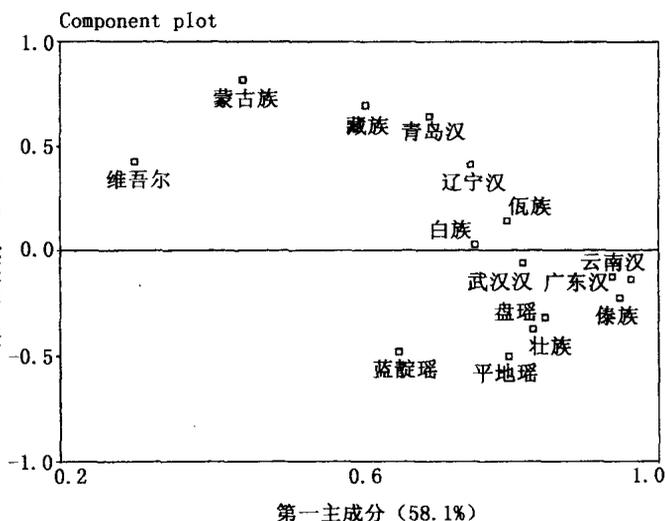


图 1 线粒体单倍群频率主成分坐标图  
(第一第二主成分共占总方差 76.4%)

Fig. 1 Principal component analysis coordinate distribution picture based on mtDNA haplogroup frequency

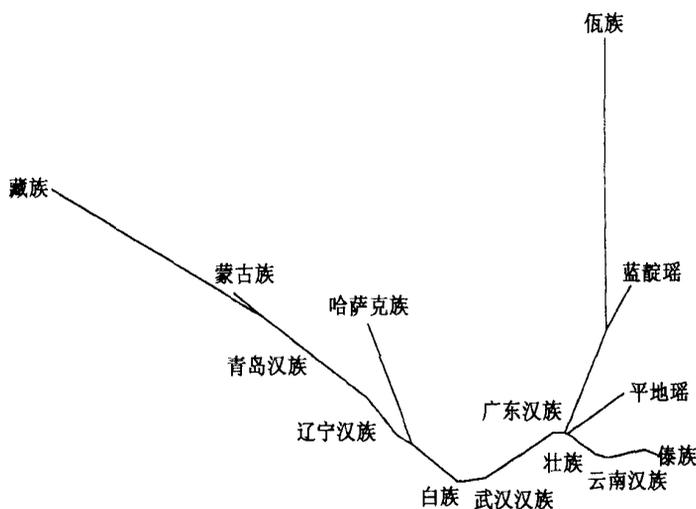


图 2 计算高变 I 区序列  $d_A$  遗传距离矩阵所得的 ME 法分子系统树

Fig. 2 ME phylogenetic tree based on  $d_A$  matrix

### 3 讨 论

在 mtDNA 单倍群上, B, F, M7 是瑶族群体主要单倍群, 分别共占盘瑶 63%, 占蓝靛瑶 60%, 占平地瑶 63%, 构成了这三个瑶族群体的主体部分。B、F、M7 在东亚呈现从南至北递减的趋势, 属于南方特征单倍群。而在东亚北方民族中较高频的 D、A、M8、M9、G、N9 等<sup>[4]</sup>则在瑶族群体中只占较少的比例甚至没有。这些提示了瑶族在线粒体遗传结构上属于典型的南方群体。

更为清晰的分析结果可以从主成分分析和分子系统分析看到。主成分分析基于高变区、RFLP 特异位点的单倍群分类。分子系统分析基于高变 I 区序列数据。这两种不同遗传标记和分析手段的平行运用得出一致的结果, 使得结果的可靠性大大加强。

盘瑶、蓝靛瑶、平地瑶三个支系群体在线粒体遗传结构(即母系遗传结构上)相互接近, 有较强的亲缘关系。瑶族在线粒体遗传结构上属于南方群体, 和北方少数民族以及北方汉族构成中国北方群体相差较大。而与地理位置邻近的周边民族诸如壮族, 南方汉族较为接近, 遗传距离与地理间隔距离呈现一定程度的正比, 显示了他们之间存在的亲缘关系。

造成这种相似的可能原因有两种, 一是瑶族先民和壮族、傣族等南方其他民族共同起源, 在形成语言民族的差异之前属于同一原始集群, 和中亚或北亚起源的北方原始集群相对, 在较早的时期分离开。目前观察到的母系遗传结构近似现象是南方共同起源的遗留。第二种可能性是瑶族和壮族, 南方汉族母系遗传结构近似是在较晚近的历史时期, 随着瑶族的不断迁徙, 和当地其他民族相互通婚融合, 进行遗传交流的结果。目前的材料倾向于支持后一种假设。按照古史传说记载, 瑶族先民所属的三苗集团在和汉人先民的炎黄集团争斗失败后, 从北方逐步迁徙到中国南方一带。传说遥远不可考证, 确认为史实的是在汉代瑶族先民居住在湘西武陵一带, 大规模迁徙入广西的时间, 据瑶族自述族谱如《千家洞流水帐》记载是在元以后, 在明清时移入广西<sup>[9]</sup>, 按照史学考证最早也在入宋以后。因此瑶族定居在广西是十分晚近的事件。在此期间, 瑶族和当地其他民族(主要是壮族和当地汉族, 而当地汉族的血统里同样含有数量可观的壮族成分)相互融合通婚的记载有很多, “瑶亦有数种, 有熟瑶, 有生瑶。生瑶在穷谷中不与华通, 熟瑶与州民同化日久, 自忘其为瑶族壮族, 外人难以分辨”<sup>[9]</sup>。平地瑶就是最早被同化和外族通婚的瑶族之一。金秀县曾有过壮族人和瑶族混居而最终融入瑶族的例子。大量的史实证明, 广西瑶族在历史上和当地其他民族有过规模较大的基因交流, 这与本实验的结果相互印证。

#### 参考文献:

- [1] Wallace D C, Brown M D, Lott M T, *et al.* Mitochondrial DNA variation in human evolution and disease[J]. *Gene*, 1999, **238**(1): 213-230.
- [2] Ingman M, Kaessmann H, Paabo S, *et al.* Mitochondrial genome variation and the origin of modern humans [J]. *Nature*, 2000, **408**(6813): 708-713.
- [3] Kong Q P, Yao Y G, Sun C, *et al.* Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences[J]. *Am J Hum Genet*, 2003, **73**: 671-676.
- [4] Yao Y G, Kong Q P, Bandelt H J, *et al.* Phylogeographic differentiation of mitochondrial DNA in Han Chinese [J]. *Am J Hum Genet*, 2002, **70**(3): 635-651.
- [5] Macaulay V, Richards M, Sykes B, *et al.* Mitochondrial DNA recombination-no need to panic[J]. *Proc R Soc Lond B Biol Sci*, 1999, **266**(1433): 2037-2039.
- [6] Kivisild T, Bamshad M J, Kaldma K, *et al.* Deep common ancestry of indian and western Eurasian mitochondrial DNA lineages[J]. *Curr Biol*, 1999, **9**(22): 1331-1334.
- [7] Yao Y G, Nie L, Hapending H, Fu Y X, *et al.* Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity[J]. *Am J Phys Anthropol*, 2002, **118**(1): 63-76.
- [8] 文波, 石宏, 任玲, 等. Y染色体、线粒体 DNA 多态性与云南宁蒨摩梭人的族源研究[J]. 中国科学

C 辑, 2003, 33(4): 375-384.

- [9] 刘祥学. 明清时期桂东北地区瑶族及其与其他民族的相互影响[J]. 中央民族大学学报(社会科学版), 1999, (5): 81-91.

## Analysis of mtDNA Polymorphisms among Three Mien Populations in Guangxi

Gao Song<sup>1</sup>, Mao Xianyun<sup>1</sup>, Wen Bo<sup>1</sup>, Li Hui<sup>1</sup>, Song Xiufeng<sup>1</sup>,  
Huang Weirju<sup>2</sup>, Zhang Youjun<sup>2</sup>, Jin Jianzhong<sup>1</sup>, Jin Li<sup>1</sup>

(1. Center for Anthropological Studies, School of Life Sciences, Fudan University, Shanghai 200433, China;

2. Institute of Ethnology and Anthropology, Guangxi University for Nationalities, Nanning 530006, China)

**Abstract:** Owning a great number of branches, Mien is an official nationality of Hmong-Mien colony, which is broadly distributed in Southwest China. With the method of HVS-1 sequencing and RFLP, 100 samples belonging to three branches of Mien were genotyped and compared with several populations spreaded in North and South China, all those analyses are presented in this article. The conclusion is that the mtDNA genetic structure of Mien, which resembles the ones of its neighborhood such as Daic, makes it belong to South China cluster.

**Keywords:** Mien; mtDNA; polymorphism; Iu Mien; Kim Mun; Lowland Mien

## 中国城市生态建设论坛 2004 年会

2004 年 6 月 12 日~13 日中国城市生态建设论坛在崇明东滩国际会议中心拉开帷幕。本论坛由复旦大学、中国生态学会城市生态专业委员会、上海市生态学会和上海市崇明县人民政府共同承办。中国科学院蒋有绪院士、中国工程院唐孝炎院士、李文华院士、安徽省人大常委会副主任/中国生态经济学会副理事长季昆森、中国生态学会副理事长王如松教授、周纪伦教授、肖笃宁教授、吴人坚教授等生态学知名专家；崇明县人民政府县长孙雷等县领导，复旦大学校党委副书记陈立民以及环境科学与工程系系领导王德耀、陈建民出席了本次论坛。城市生态研究中心主任王祥荣教授作为论坛执行主席主持会议。12 日上午，论坛主题报告会由唐孝炎院士主持。专家学者们从自然、经济、社会、人文等角度论述了生态城市建设、规划与管理及崇明生态岛域总体规划建设。蒋有绪院士认为：“21 世纪是人类纠正生态恶化最关键的 100 年”，他概括论述了生态城市建设中科学的自然观与资源观，强调人与自然、人口资源与环境的协调。李文华院士从城市发展历程伴随的生态环境恶化出发，提出生态城市建设值得思考的问题及生态城市建设的框架。王如松教授、季昆森研究员分别就“城市生态规划、管理”及“循环经济与生态型城镇建设”提出了全新的观点。崇明县人民政府副县长胡俊教授就崇明生态岛域总体规划建设情况向与会专家、教授做了汇报。13 日讨论，会后部分著名专家与崇明政府领导进行了电视专场对话并提出了具有建设性意义的意见及建设。（由王祥荣教授供稿）