

浒苔小 RNA 高通量测序及相关生物信息学分析

黄爱优^{①②}, 王广策^{①*}, 何林文^{①②}, 牛建峰^①, 张宝玉^①

① 中国科学院海洋研究所, 青岛 266071;

② 中国科学院研究生院, 北京 100049

* 联系人, E-mail: gcwang@qdio.ac.cn

2011-04-08 收稿, 2011-07-04 接受

国家科技支撑计划(2008BAC49B01)和国家自然科学基金(30830015)资助项目

摘要 浒苔有定生和漂浮两种状态, 两者的形态结构没有显著差别, 生长速度却有很大差异, 推测浒苔可能存在重要的增殖调控模式. **microRNA** 是非常重要的转录后调控因子, 已证明其普遍存在于真核生物, 在生长发育等过程中起重要作用. **miRNA** 在高等动植物中研究的比较多, 而在藻类里则相对较少. 为了研究浒苔的 **miRNA** 及其在漂浮浒苔快速增殖中的可能作用, 采用 **solexa** 测序的方法对浒苔小 RNA 进行测序并做了相关的生物信息学分析, 结果表明, 浒苔中存在丰富的小 RNA, 它们可能在浒苔各种生物过程中起重要作用.

关键词

浒苔
小 RNA, **miRNA**
solexa 测序
生物信息学

浒苔(*Ulva prolifera*)是绿藻门(Chlorophyta)绿藻纲(Chlorophyceae)石莼目(Ulvales)石莼科(Ulvaceae)石莼属(*Ulva*)绿藻的总称^[1,2]. 浒苔有定生和漂浮两种状态. 定生浒苔常固着在潮间带的岩石、沙砾上, 以正常速度生长; 定生浒苔断裂后可随海流漂浮, 形成漂浮浒苔. 漂浮浒苔生长迅速^[3], 在短时间内(24 h)可以形成巨大的生物量, 形成“绿潮”^[4], 对海洋生态尤其是海水养殖和观光旅游等产业产生重大负面影响. 尽管两者的形态结构没有显著差别, 但定生浒苔生长速度正常, 漂浮浒苔的增殖速度异常, 由此推测浒苔可能存在重要的增殖调控模式.

microRNA (miRNA) 是一类长约 21 nt 的内源单链非编码小 RNA 分子, 是非常重要的转录后调控因子^[5-7]. **miRNA** 广泛存在于真核生物中, 并在相近的物种里存在高度的保守性^[8,9]. **miRNA** 通过互补配对与 3'非翻译区(多见于动物)或基因编码区(多见于植物)相互作用, 引起翻译抑制(多见于动物)或靶基因的降解(多见于植物)^[8,10], 研究发现其在个体发育调控、细胞分化、逆境应答等生物过程中起重要作用^[8].

传统的 **miRNA** 研究方法有直接克隆法、生物信息学预测方法等. 传统直接克隆法的缺点是只能分

离高丰度表达的 **miRNA**, 难以发现一些表达量低或只在特定阶段表达的 **miRNA**^[11,12]. 生物信息学预测方法对于 **miRNA** 的表达量、表达与否都无要求, 能发现一些低量表达的 **miRNA**, 但其缺点是假阳性太高, 必须用大量的实验进行验证^[13,14]. 最近发展起来的结合高通量测序与生物信息学分析于一体的 **miRNA** 研究方法, 能有效、快速地分离鉴定 **miRNA**. 其基本思想是, 从总 RNA 里分离出小 RNA, 加上 5'接头和 3'接头, RT-PCR 扩增后直接进行高通量测序, 然后用生物信息分析方法对所得的小 RNA 序列进行分析^[15-18]. **miRNA** 在高等动植物中研究的比较多, 在藻类里则研究的比较少. 为了研究浒苔的 **miRNA** 及其在漂浮浒苔快速增殖中的可能作用, 我们从浒苔里分离并建立小 RNA 文库, 采用 **solexa** 测序的方法对其进行测序并做了相关的生物信息学分析, 以期分析浒苔增殖的分子调控机制.

1 材料与方法

(i) 样品处理. 新鲜浒苔丝状体采自青岛海边, 蒸馏水洗涤干净后, 用吸水纸吸去表面的水分, 液氮速冻后存于-80℃待用.

英文版见: Huang A Y, Wang G C, He L W, et al. Characterization of small RNAs from *Ulva prolifera* by high-throughput sequencing and bioinformatics analysis. Chinese Sci Bull, 2011, 56, doi: 10.1007/s11434-011-4678-6

(ii) RNA 提取. 样品在液氮中研磨至粉末状, 加入 Trizol 试剂混合均匀, 按照 Trizol 试剂说明书提取总 RNA.

(iii) 小 RNA 文库的构建及测序. 总 RNA 用 15% 变性聚丙烯酰胺凝胶电泳分离后, 回收纯化 18~28 nt 的部分, 在 T₄ RNA 连接酶的作用下加上 5' 和 3' 接头, 所得产物经 RT-PCR 扩增之后用 solexa 高通量测序方法进行测序.

(iv) 序列基本分析. Solexa 测序所得的序列经如下程序进行分析(图 1): (1) 去接头、去低质量序列和去污染处理, 得到高质量序列, 对其进行长度分布统计; (2) 去掉小于 18 nt 的序列, 得到纯净序列; (3) 利用 SOAP 将剩下的序列比对到浒苔 EST 序列及莱茵衣藻、拟南芥、三角褐指藻基因组, 分析小 RNA 在浒苔 EST 序列上及这些模式生物基因组上的分布特征; (4) 将小 RNA 序列与 GenBank 数据库及 Sanger 中心的 Rfam 数据库中的 ncRNA (非编码 RNA) 分别

进行比对, 识别小 RNA 序列中的 ncRNA 的降解片段; (5) 将 clean reads 中的片段互相比对, 找出可以互补、符合 siRNA 结构特征的成对小 RNA, 作为可能的 siRNA 候选者; (6) 通过与 miRBase 数据库中的所有植物 miRNA 进行比对, 识别其中的 miRNA 序列, 小 RNA 片段与已知植物 miRNA 序列的相似性在 90% (不多于 2 个错配) 以上的序列认为是 miRNA 序列; (7) 统计各已知 miRNA 家族序列的表达数, 构建已知 miRNA 表达谱; (8) 将小 RNA 序列与拟南芥基因组中的重复序列进行比对, 识别来自重复序列的小 RNA 片段; (9) 将小 RNA 序列与拟南芥的 mRNA 进行比对, 以识别其中的 mRNA 降解片段. 完成以上的比对分析以后, 我们将所有 sRNA 与各类 RNA 的比对结果进行总结, 以对所有小 RNA 进行注释. 在以上各注释信息中, 若存在一个 sRNA 同时比对上 2 种不同的注释信息的情况, 则按照 rRNA 等 > 已知 miRNA > 重复序列 > 外显子 > 内含子的优先级顺

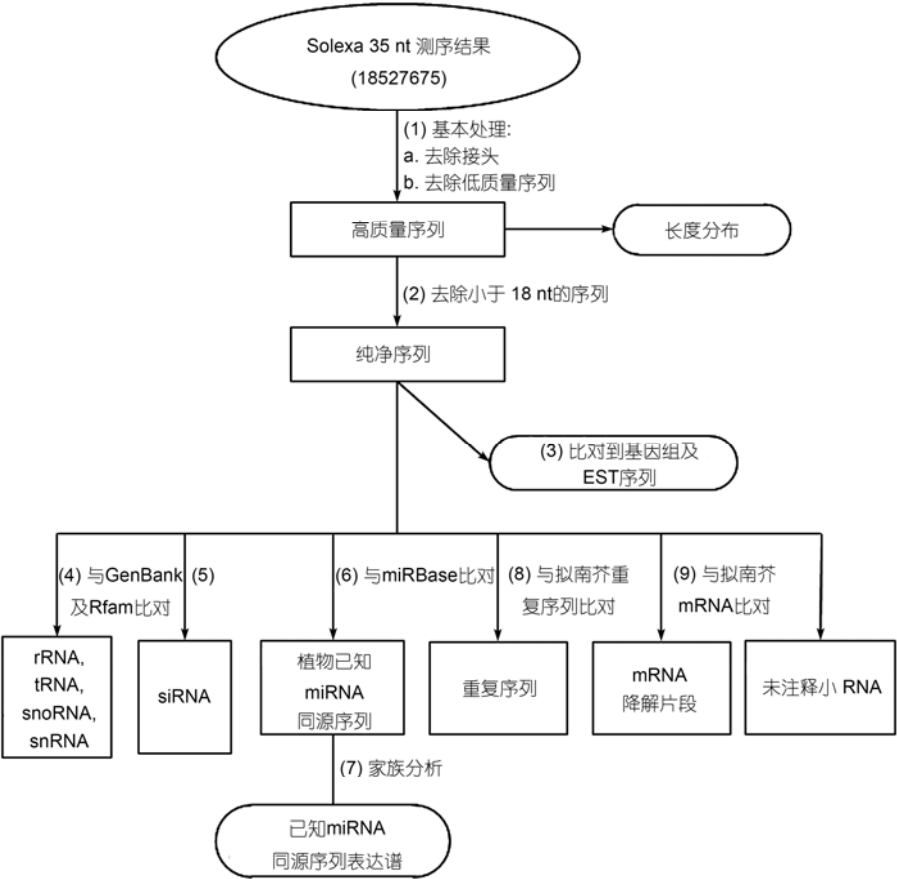


图 1 浒苔小 RNA 基本生物信息分析流程图

序将 sRNA 注释^[19], 没有比上任何注释信息的 sRNA 用 unann 表示. 由于 rRNA 等是由 GenBank 和 Rfam 两个数据库比对所得, 规定这 2 个数据库间的优先级为 GenBank > Rfam.

(v) 新 miRNA 的预测. 我们选取拟南芥的基因组及浒苔的 EST 序列进行新 miRNA 预测的参考序列. 未注释的小 RNA 比对到参考序列后, 截取其周围上下游各 300 nt 的序列, 探寻其二级结构及 Dicer 酶切位点信息、能量等特征, 以寻找新 miRNA. 所用的筛选参数如下: miRNA 在基因组上完全匹配的位点 ≤ 20 , 前体自由能小于或等于 -18 kcal/mol (Mfold), miRNA 与 miRNA* 之间的长度小于或等于 300 nt, miRNA 与 miRNA* 之间有大于或等于 16 个碱基对的配对, 不多于 4 个连续突起或错配.

(vi) 新 miRNA 靶基因预测. 我们用 miRanda 来预测新 miRNA 的靶基因. 所用到的参数如下: 匹配分数 $S \geq 90$, 自由能 $\Delta G \leq -20$ kcal/mol; scaling parameter = 2. 所得的 miRNA-靶基因配对再按 Allen 等人^[12]建议的标准进行手动筛选. 简而言之, miRNA 与靶基因配对的第 2~21(自 miRNA 5'端数起)位不超过 4 个错配. 第 2~12 位无连续错配, 所有位点无连续 2 个以上错配, 第 10~11 位无错配, 第 1~12 位不多于 2.5 个错配. G:U 配对看成 0.5 个错配. miRNA 与靶基因配对的自由能必须大于其完美配对时自由能的 74%.

(vii) RT-PCR 验证部分已知 miRNA 的表达. 我们用 RT-PCR 的方法对部分 miRNA 的表达进行了验证. 具体步骤为, 总 RNA 加上多聚腺苷酸尾巴之后, 用加接头的(poly-T-adapter)引物进行反转录, 然后用根据成熟 miRNA 的序列设计的正向引物和接头进行 PCR 扩增, 目的片段在 60~80 bp 之间.

2 结果

2.1 浒苔小 RNA

为了寻找浒苔的 miRNA, 我们构建了浒苔小 RNA 文库并用 solexa 高通量测序方法进行了测序. 去除接头并滤去低质量数据后, 我们得到了长度分布在 10~43 nt 的大量浒苔小 RNA 序列, 其中以 25 nt 的序列居多(图 2). 去掉小于 18 nt 的序列后, 我们最终得到了 16052321 条纯净序列(表 1), 其中特异序列的数目为 3460994. 我们用 SOAP 将这些小 RNA 与

浒苔 EST, 拟南芥、莱茵衣藻、三角褐指藻的基因组进行比对, 其中只有 1338946 (8.34%)的总序列及 72493 (2.09%)的特异序列能匹配(图 3, 表 2), 其中大部分匹配的是浒苔 EST, 而与其他物种的基因组则很少匹配. 有不少小 RNA 与拟南芥基因组重复序列匹配, 其中与 LSU-rRNA_Ath: 1 及 SSU- rRNA_Ath: 1 比对的上的明显多于其他重复序列(表 S1). 我们将浒苔小 RNA 与 GenBank, Rfam, miRBase 进行了比对, 并将小 RNA 序列两两比对以寻找 siRNA, 最后进行了注释(图 4, 表 3). 由于缺少全基因组序列, 约 92% 的小 RNA 序列未能注释. 所注释的小 RNA 里, rRNA 最多, 占总数的 3.64%, 其次是 siRNA, 占总数的 1.96%, tRNA 占 1.90%. 其余的都只占很少一部分.

2.2 已知 miRNA 表达谱的构建及进化分析

为了寻找浒苔中的已知 miRNA, 我们将浒苔小 RNA 与 miRBase 中的所有植物 miRNA 进行了比对. 16052321 的总序列里, 有 20657 的特异序列与植物 miRNA 有同源性(表 3). 在允许 1~2 个错配的情况下, 这些 miRNA 分属于 285 个已知的 miRNA 家族(表 S2). 这些家族的 miRNA 表达量从 1 个到上万个拷贝不等

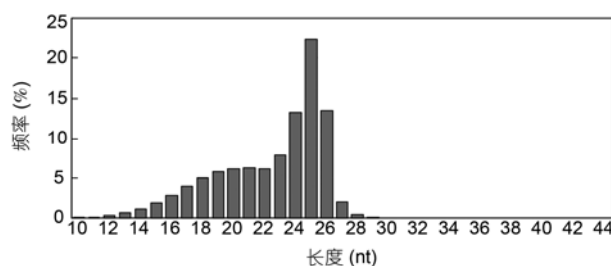


图 2 浒苔小 RNA 长度分布图

表 1 浒苔小 RNA 测序数据处理及片段长度分布统计

片段类型	数量	百分比(%)
总片段数	18527675	
高质量片段数	18197647	100
3' 接头	55009	0.30
插入片段	13993	0.08
5' 接头	77201	0.42
小于 18 nt 的片段数	1998394	10.98
polyA	729	0.00
纯净序列数	16052321	88.21

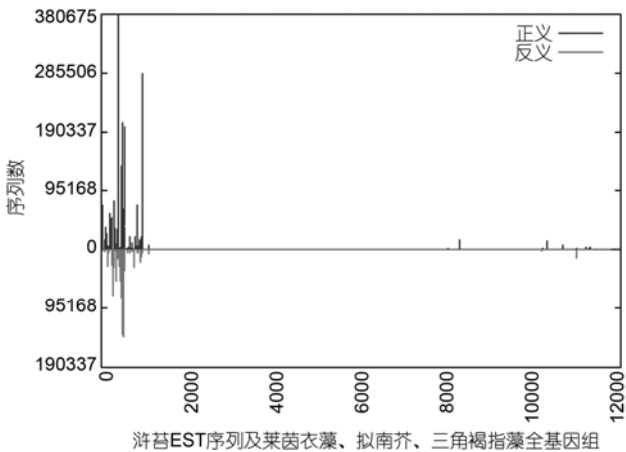


图 3 浒苔小 RNA 在其 EST 序列及莱茵衣藻、拟南芥、三角褐指藻 3 种模式生物基因组上的分布情况
横坐标从左至右依次为浒苔 EST 序列、莱茵衣藻基因组、拟南芥基因组及三角褐指藻基因组

表 2 浒苔小 RNA 在浒苔 EST 序列及模式生物基因组上的分布情况

	特异小 RNA 数	百分比 (%)	总小 RNA 数	百分比 (%)
小 RNA 总量	3460994	100	16052321	100
比对上 EST 及基因组部分	72493	2.09	1338946	8.34

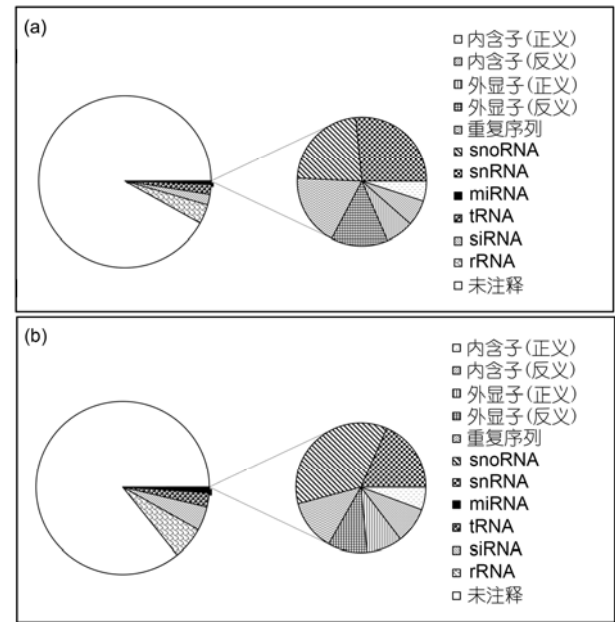


图 4 浒苔小 RNA 注释结果
(a) 特异序列注释结果; (b) 总序列注释结果

表 3 浒苔小 RNA 注释结果

分类	特异小 RNA 数	百分比 (%)	总小 RNA 数	百分比 (%)
内含子(正义)	75	0.00	240	0.00
内含子(反义)	97	0.00	379	0.00
外显子(反义)	105	0.00	377	0.00
外显子(正义)	213	0.01	415	0.00
重复序列	270	0.01	536	0.00
snoRNA	341	0.01	1515	0.01
snRNA	398	0.01	784	0.00
miRNA	20657	0.60	170863	1.06
tRNA	65680	1.90	420839	2.62
siRNA	67866	1.96	710322	4.43
rRNA	125862	3.64	974059	6.07
未注释小 RNA	3179430	91.86	13771992	85.79
总数	3460994	100	16052321	100

(图 5). 其中以 miR1520 的表达量为最高, 达 15489 个拷贝, miR2678 及 miR1874 的表达量也都超过了 10000 个拷贝.

2.3 新 miRNA 的预测

miRNA 前体序列能形成稳定的发夹结构, 这为新 miRNA 的预测提供了有力的依据. 由于浒苔全基因组未测序, 我们只好用拟南芥的全基因组及浒苔 EST 序列来预测新的 miRNA, 但没有预测到相应的新 miRNA.

2.4 部分已知 miRNA 的实验验证

我们用 RT-PCR 的方法对测序丰度最高的 10 个已知 miRNA 的表达进行了验证. RT-PCR 的结果有 6 个 miRNA 有约 60 bp 的目的带, 有的还有大些的带, 可能是相应的 miRNA 前体(图 6). 这充分说明了这些

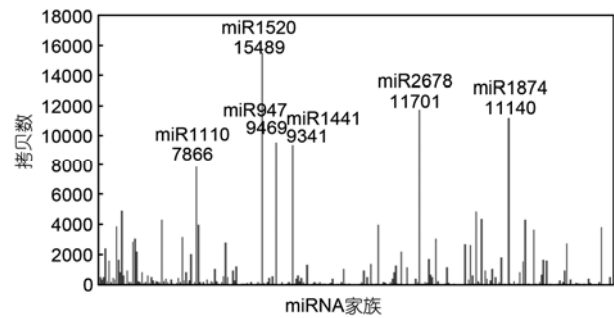


图 5 浒苔中已知 miRNA 的表达丰度

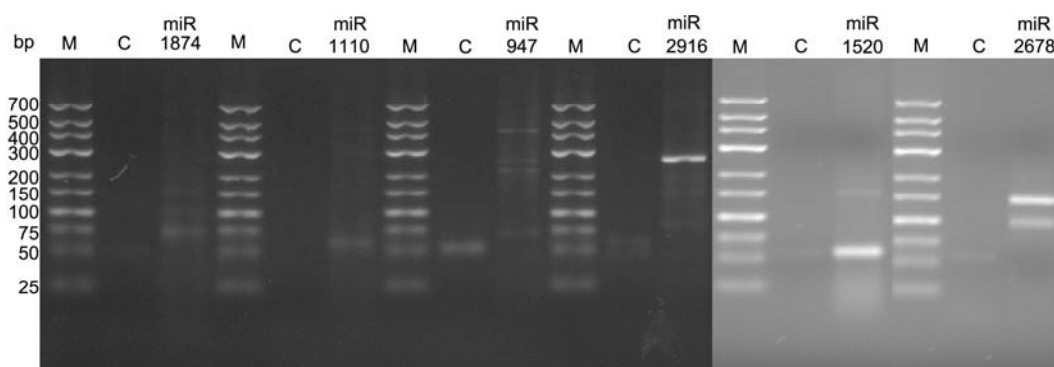


图 6 浒苔中部分已知 miRNA 的实验验证

M, Marker; C, 对照

miRNA 在浒苔里确实存在并表达。

3 讨论

实验结果得到了 16052321 条纯净序列(表 1), 代表 3460994 条特异序列, 表明浒苔中可能存在复杂的小分子 RNA 调控机制。测序得到的序列与拟南芥、莱茵衣藻及三角褐指藻这 3 种模式生物的基因组比对, 只有 8.34% 的总序列及 2.09% 的特异序列匹配, 说明浒苔小分子 RNA 调控机制可能有其不同于这些模式生物的特殊性。有不少小 RNA 与拟南芥基因组重复序列比对得上, 说明浒苔小 RNA 分子可能在干扰重复序列中起到重要作用。小 RNA 注释结果显示, siRNA 数目众多, 在已注释的小 RNA 分子中仅次于 rRNA, 表明浒苔中可能存在复杂的 siRNA 沉默机制。

浒苔小 RNA 分子长度最多的为 25 nt, 这与拟南芥中最多的为 24 nt 的情况不一样^[12,15]。据报道, 小立碗藓及莱茵衣藻也都缺少 24 nt 的小 RNA^[11,20,21]。小 RNA 的长度与加工其生成的酶的种类有关, 如经 DCL2 加工的小 RNA 长度为 24 nt, DCL1 加工生成的小 RNA 长度为 21 nt^[20]。浒苔、小立碗藓及莱茵衣藻的小 RNA 的长度与拟南芥不同, 说明这些进化上相对低等的光合生物参与小 RNA 加工的酶的种类和丰度可能有别于拟南芥。事实上, 我们根据拟南芥、小立碗藓、水稻等 9 个物种的 Dicer-like 酶设计了简并引物, 以浒苔 cDNA 为模板进行 PCR 扩增, 得到的带与预期的相符, 测序结果表明, 其与其他物种的 DCL 确实存在一定的序列相似性, 说明浒苔中确实存在 miRNA 加工所必需的酶。至于这些酶在浒苔中具体怎么起作用, 还需进一步的实验验证。

浒苔 sRNA 中有部分和植物已知 miRNA 有高度同源性, 相似度在 90% 以上, 表明浒苔的 miRNA 调控方式可能与植物有一定的相似性。其中有些 miRNA 家族表达量特别高, 如 miR1520, miR2678 及 miR1874 的表达量都超过了 10000 个拷贝, 它们可能在浒苔中起重要调控作用。为了进一步验证这些 miRNA 在浒苔中的表达, 我们用 RT-PCR 的方法对测序丰度最高的 miRNA 进行了实验验证, 其中有 6 个得到了验证, 说明这些 miRNA 确实在浒苔中存在并表达。我们尝试将扩出的 miRNA 片段进行回收测序, 但是可能由于目的片段很弱, 且片段太小(仅几十 bp), 而胶回收试剂盒对小片段的回收效率非常低, 因此我们没能成功回收。我们也试图分析这些 miRNA 在浒苔中的功能, 但由于缺乏相关的基因组信息, 本研究只分析这些 miRNA 家族在其他物种中的作用。这些 miRNA 都是最近发现的, 功能也尚未明确。例如, miR1520, miR2678 及 miR1531 是在大豆的根部发现的, 可能与根瘤菌的共栖有一定的关系^[22,23], 我们也猜测它们可能在根部的快速增殖中起作用。另外 miR947 在松树里可能与其抗旱有关^[24], 而我们前期的工作也表明浒苔在干出至含水量仅为正常情况的 20% 的时候仍有微弱的光合活性, 在复苏 30 min 后光合活性完全恢复正常^[25], 说明其耐旱性很强, 我们猜测 miR947 可能也与浒苔的耐旱有关, 但这些都需进一步的实验验证。浒苔 sRNA 中还有约 92% 的序列未能注释, 可能是因其与其他光合生物亲缘关系比较远, 可能存在特有的 miRNA。但由于缺乏基因组信息, 我们没能对这些可能的已知 miRNA 进行前体验证, 也没能预测新的 miRNA。为了弥补这一缺憾, 我们用浒苔的 EST 进行了前体验

证的工作, 但仍然没有得到有用的结果, 可能原因是 EST 数据太少(仅 1000 多条).

根据靶基因预测的结果我们得到了 4 个潜在的

靶基因位点. 至于 miRNA 在浒苔中到底起到什么样的作用, 是否在定生浒苔及漂浮浒苔的生长发育过程中起调控作用, 还需要进行进一步的验证.

参考文献

- 1 曾呈奎, 张德瑞, 张峻甫, 等. 中国经济海藻志. 北京: 科学出版社, 1962
- 2 Hayden H, Blomster J, Maggs C, et al. Linnaeus was right all along: *Ulva* and *enteromorpha* are not distinct genera. *Eur J Phycol*, 2003, 38: 277–294
- 3 Gao S, Chen X Y, Yi Q Q, et al. A strategy for the proliferation of *Ulva prolifera*, main causative species of green tides, with formation of sporangia by fragmentation. *PLoS One*, 2010, 5: e8571
- 4 Fletcher R. The occurrence of “Green Tides” – a review. In: Schramm W, Nienhuis P H, eds. *Ecological Studies*, Vol. 123. Marine Benthic Vegetation. Recent Changes and the Effects of Eutrophication. Berlin: Springer, 1996. 7–43
- 5 Lau N, Lim L, Weinstein E, et al. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, 2001, 294: 858
- 6 Lee R C, Ambros V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*, 2001, 294: 862–864
- 7 Lagos-quintana M, Rauhut R, Lendeckel W, et al. Identification of novel genes coding for small expressed RNAs. *Science*, 2001, 294: 853
- 8 Bartel D P. MicroRNAs genomics, biogenesis, mechanism, and function. *Cell*, 2004, 116: 281–297
- 9 Reinhart B J, Weinstein E G, Rhoades M W, et al. MicroRNAs in Plants. United State: Patent Application Publication, 2004
- 10 Millar A, Waterhouse P. Plant and animal microRNAs: Similarities and differences. *Funct Integr Genomics*, 2005, 5: 129–135
- 11 Arazi T, Talmor-neiman M, Stav R, et al. Cloning and characterization of micro-RNAs from moss. *Plant J*, 2005, 43: 837–848
- 12 Xie Z, Allen E, Fahlgren N, et al. Expression of *Arabidopsis* miRNA genes. *Plant Physiol*, 2005, 138: 2145
- 13 Jones-rhoades M W, Bartel D P. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol Cell*, 2004, 14: 787–799
- 14 Li L, Xu J, Yang D, et al. Computational approaches for microRNA studies: A review. *Mamm Genome*, 2010, 21: 1–12
- 15 Rajagopalan R, Vaucheret H, Trejo J, et al. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev*, 2006, 20: 3407
- 16 Fahlgren N, Howell M, Kasschau K, et al. High-throughput sequencing of *Arabidopsis* microRNAs: Evidence for frequent birth and death of miRNA genes. *PLoS One*, 2007, 2: e219
- 17 Sunkar R, Zhou X, Zheng Y, et al. Identification of novel and candidate miRNAs in rice by high throughput sequencing. *BMC Plant Biol*, 2008, 8: 25
- 18 Wei B, Cai T, Zhang R, et al. Novel microRNAs uncovered by deep sequencing of small RNA transcriptomes in bread wheat (*Triticum aestivum* L.) and *Brachypodium distachyon* (L.) beauv. *Funct Integr Genomics*, 2009, 9: 499–511
- 19 Calabrese J, Seila A, Yeo G, et al. RNA sequence analysis defines dicer’s role in mouse embryonic stem cells. *Proc Natl Acad Sci USA*, 2007, 104: 18097
- 20 Zhao T, Li G, Mi S, et al. A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. *Genes Dev*, 2007, 21: 1190
- 21 Axtell M, Snyder J, Bartel D. Common functions for diverse small RNAs of land plants. *Plant Cell Online*, 2007, 19: 1750
- 22 Subramanian S, Fu Y, Sunkar R, et al. Novel and nodulation-regulated microRNAs in soybean roots. *BMC Genomics*, 2008, 9: 160
- 23 Lelandais-briere C, Naya L, Sallet E, et al. Genome-wide medicago truncatula small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. *Plant Cell Online*, 2009, 21: 2780
- 24 Lu S, Sun Y H, Amerson H, et al. MicroRNAs in loblolly pine (*Pinus taeda* L.) and their association with fusiform rust gall development. *Plant J*, 2007, 51: 1077–1098
- 25 Gao S S S, Wang G C, Niu J F, et al. Psi-driven cyclic electron flow allows intertidal macro-algae *Ulva* sp. (chlorophyta) to survive in desiccated conditions. *Plant Cell Physiol*, 2011, doi: 10.1093/pcp/PCR038

补充材料

表 S1 浒苔小 RNA 与拟南芥重复序列比对结果统计

表 S2 已知 miRNA 表达谱

本文的以上补充材料见网络版 csb.scichina.com. 补充材料为作者提供的原始数据, 作者对其学术质量和内容负责.



表 S1 浒苔小 RNA 与拟南芥重复序列比对结果统计

重复类型	特异小 RNA 数	总小 RNA 数	重复类型	特异小 RNA 数	总小 RNA 数
(ATG)n: 0	2	2	ATLINE1_6: 0	1	1
(ATG)n: 1	11	12	ATLINE1_7: 1	1	1
(CA)n: 0	2	2	ATLINE1_8: 0	2	8
(CAA)n: 0	1	1	ATLINE1_9: 0	1	1
(CAA)n: 1	1	1	ATLINEIII: 1	8	10
(CACCAT)n: 0	1	1	ATMU2: 1	4	12
(CAGA)n: 0	2	2	ATMU3: 1	1	1
(CAGA)n: 1	7	9	ATREP11: 0	1	2
(CAGAGA)n: 1	1	1	ATREP15: 1	1	1
(CAT)n: 0	7	7	ATREP18: 1	3	3
(CATCC)n: 0	1	1	ATREP1: 1	1	4
(CGA)n: 0	1	1	ATREP2A: 0	1	1
(GAA)n: 1	2	2	ATREP5: 1	1	1
(GGA)n: 1	5	5	ATREPX1: 0	1	2
(GTTTG)n: 1	1	2	ATREPX1: 1	1	1
(TA)n: 1	5	10	ATSAB5: 0	1	1
(TAGA)n: 0	1	1	AT_rich: 0	4	5
(TC)n: 0	1	1	AT_rich: 1	5	5
(TCTA)n: 0	3	9	Arnold4: 0	2	2
(TG)n: 1	2	2	ArnoldY1: 0	1	1
(TGG)n: 1	1	1	ArnoldY1: 1	4	4
(TTATA)n: 0	1	1	AtSB2: 1	15	181
(TTC)n: 0	5	5	AtSB4: 0	1	1
(TTTTTC)n: 0	1	1	Athila2_I: 0	2	2
A-rich: 1	1	1	Athila3_I: 0	1	1
ATCOPIA69A_I: 0	1	1	Athila3_I: 1	2	2
ATCOPIN_LTR: 0	4	4	Athila4A_I: 1	1	1
ATCopia17I: 1	1	1	Athila4B_LTR: 0	1	1
ATCopia19I: 1	1	1	Athila4B_LTR: 1	2	5
ATCopia21I: 1	2	2	Athila4C_I: 1	1	1
ATCopia24I: 0	3	3	Athila4D_LTR: 1	2	2
ATCopia24I: 1	1	1	Athila4_I: 0	1	1
ATCopia38B_I: 0	2	4	Athila4_I: 1	1	1
ATCopia41LTR: 1	1	1	Athila6A_LTR: 0	2	2
ATCopia56_I: 1	1	1	Athila6B_I: 0	1	1
ATCopia5I: 0	1	7	Athila6B_I: 1	2	2
ATCopia66_I: 1	1	1	Atlantys1_I: 1	1	5
ATCopia69_I: 1	1	1	Atlantys2_I: 1	5	14
ATCopia7I: 1	3	7	Atlantys3_I: 0	1	1

转下页

续表

重复类型	特异小 RNA 数	总小 RNA 数	重复类型	特异小 RNA 数	总小 RNA 数
ATCopia80_I: 0	1	1	CT-rich: 0	11	21
ATDNAI27T9A: 1	1	1	GA-rich: 0	1	1
ATENSPM1: 1	1	1	GA-rich: 1	8	8
ATENSPM3: 1	2	2	Helitron1: 0	1	4
ATENSPM6: 0	1	93	Helitron3: 1	1	1
ATENSPM6: 1	1	1	Helitron4: 0	1	1
ATENSPM9: 0	2	3	Helitron5: 1	1	1
ATENSPM9: 1	1	1	HelitronY1D: 0	1	5
ATGP10LTR: 1	3	14	LSU-rRNA_Ath: 0	120	414
ATGP1I: 0	1	1	LSU-rRNA_Ath: 1	6450	191463
ATGP1I: 1	2	2	SSU-rRNA_Ath: 0	84	288
ATGP1LTR: 0	1	1	SSU-rRNA_Ath: 1	4268	95415
ATGP3I: 0	6	8	Sadhu7-2: 1	1	1
ATGP5I: 1	1	1	TA11: 1	1	2
ATHAT1: 1	2	2	TNAT2A: 0	1	1
ATHATN1: 0	1	1	Vandal11: 0	1	1
ATHILA6C_I: 0	1	1	Vandal1: 0	2	2
ATHPOGON1: 1	1	1	Vandal20: 1	2	3
ATIS112A: 1	1	1	Vandal21: 0	2	2
ATLINE1A: 0	2	4	Vandal5: 0	1	1
ATLINE1_10: 0	1	1	Vandal5: 1	1	1
ATLINE1_11: 1	2	2	Vandal8: 1	1	1
ATLINE1_1: 0	1	1	ambi	62	158
ATLINE1_1: 1	1	1	total	11237	288349
ATLINE1_4: 1	2	2			

表 S2 已知 miRNA 表达谱

miRNA 家族	表达量	序列(5'→3')	miRNA 家族	表达量	序列(5'→3')
miR1023	1122	AGAGAACTCGGGTGGAGGCAG	miR2111	246	TATCTGCATCTGAAGTCAA
miR1024	150	TCTGGTTTGATTGTGGCT	miR2112	1	CGTAAATGGGATAATCAATGA
miR1026	1539	TGAAATGACTTGAGAGGTAGCA	miR2118	3	TGCCCAATGCCTGTGCCATGCCAAA
miR1028	332	TTTAGATCAACTATGCACC	miR2119	555	CAAAGGAGTGGTAGGGGA
miR1030	70	TATGCACTGGCACCTGCACGG	miR2123	93	TTAAAAAGTCAACTGTCAAA
miR1031	22	TCTATTCTCTGGGGCTTCT	miR2125	1	TTTTTCTATAGACTATGATCCAT
miR1033	51	TAGAGCGGGCCGTGATGGGCAT	miR2199	5	TGATACACTAGCACGGATCAC
miR1034	7	TTATCTGGCAGCGCTGATGT	miR2275	23	AGTAGTTGGAGGAAGCAAA
miR1036	98	ACTATTAAGGATCTATAC	miR2594	2	TCCATGGCTCAAGGATGACAGCG
miR1037	2184	CCTTATTGGATTGGATGG	miR2597	67	TTTGACTTGGACGATTTG
miR1038	326	TAGGTTCGTTTCCTACCTCAAG	miR2601	32	TTTTTGTAACGCTTGTTGGATCCC
miR1039	417	AGGGAACGGGCTCAGAATC	miR2604	4	AATTGTTATGTGGGAAGTGTG
miR1041	414	TCTTCGGGTGCATGCAGTGGATCCT	miR2607	50	ATATGATTATGTAGATAGGT
miR1043	1040	ATGCGCGTGAATGATGAAGGCCT	miR2610	8	ACATTGGACTTGTGATGGCT

转下页

续表

miRNA 家族	表达量	序列(5'→3')	miRNA 家族	表达量	序列(5'→3')
miR1044	1	TTGTAGTGCAGAGTCTGTTTT	miR2611	105	TATTTGTTGGTTTGAATGAA
miR1047	77	TGATTCAGCAGATGACCTGTG	miR2614	167	CGGTTCCGACTCCGGTATTC
miR1048	601	TGAACATGAGAGTGGACGA	miR2626	911	AAGGTCGGGATGAGGTGTT
miR1061	1643	TGTTAGCATGGAATAATG	miR2628	22	CACTGAAAGAAGATGAGTG
miR1069	13	TCGATATCAAAGTGCTACT	miR2630	69	GGTTTTGATTTTGGTATTT
miR1076	3	TGAACACAAGGCGCGAAATTG	miR2631	592	TGATACACCACGTGCAACT
miR1077	341	ATCAAATGTTTCGGATAGTG	miR2634	2053	TTTATTCTCGTTGGTTTC
miR1083	194	TAGCCTGGAAAGAAGACT	miR2635	23	ATTATGGTCAACGGTCAG
miR1087	13	TACAGTGACTGGTGTTAATGCTT	miR2644	151	CATCAGCTTGATGGTAGTGT
miR1089	462	AGACAGCTAGGATGTTGC	miR2645	400	TTTGTAGAGATAGCACATT
miR1090	16	AAAGCCCATTCTCAACAAGA	miR2651	533	TTTGATTGGATGACGCTATT
miR1091	4259	CGGGTGACGGAGGATTAGC	miR2652	47	AAAGCAGGGCATAAGGATT
miR1092	25	TGACTGGAATAGTGGTGTT	miR2653	11	TCACCGCTGCTGTGAATATGATTGC
miR1093	44	TGGAGGCGGTCTCGCCAAGG	miR2656	235	AAGAGCATAATCGGTAGG
miR1098	128	TGATGACGTTGTATGCCTGAAATC	miR2660	1	TAGAGACTTCAGGATAAGCTA
miR1099	7	TATATGCAGTGGAATTTTTGTC	miR2661	10	CAGGTTTGAAAAAGAGGCAG
miR1100	2	TGTACTGACAGAACACGACCACTC	miR2663	51	TGTCGAGAGGGCGTTCATT
miR1106	6	TTCAAAGGTCTGTAATGTGTG	miR2666	272	AGAAAGTGAGGTACATGCAAGGA
miR1108	495	TGACTGAGAGACATACCG	miR2667	226	TCCTTGAATCTGACGCGGTACT
miR1110	7866	GCAGGGCGGTGGTCATGGA	miR2677	13	TTTATTGATATTCTAATT
miR1113	2789	TGACAGTCCTAAGGTAGC	miR2678	11701	TGAACATTGTTGAGACGAGTGACTT
miR1114	15	AAGGGACTCAAGAACATTACAGCAAG	miR2679	1663	CGGTTTCATTTCGGAACGGGTG
miR1115	442	TGAGGTAAGAGCATTTGGTG	miR2680	1067	TCCTCGGACCTACGTGTGGT
miR1117	170	AGTACCGTTCGTGCAGGAATCC	miR2862	14	TCCAACACACTTCAGACTTCGTGCC
miR1118	42	TATACATGGTAAGGAATGGAGGGA	miR2863	192	TTCGTGATTTGGACTAAGTC
miR1119	23	TGGCACGGACGTTCTTGAGTCAG	miR2864	2	ATTGTTTGCAATTGGATAGGT
miR1120	10	AATTGCTTCATATTATGGACCGAG	miR2866	1181	TCAGTTTGTTAGAGCATC
miR1122	112	TACATACATCGTATTTGCGT	miR2875	179	ATGTCTAAGTATAAACAGTTTATA
miR1127	4287	TAGCTTTGTTTCGGAATTAC	miR2880	1349	ACGGTAATGCCGTTTCGGAAGTG
miR1131	101	TATACCGGTCGTGGGCTCACC	miR2906	25	AAGGCGGCGGCACTGCTGG
miR1132	376	CAATTATGGATCGTGAACGAG	miR2910	1624	TCTTAGTTGGTGGAGCGATTTGT
miR1134	4	AAACAACAAAAGATGGAAGAAGA	miR2912	15	TCATCAGGAACCTCGAGA-TAATGTGGC
miR1135	32	TGTGACAAGTATTCACGAACGGG	miR2914	799	GTGGTGGAACGGGTGACGGA
miR1140	891	ACGCCTACAATCAGTCGGAGC	miR2915	785	TCCCCTGTAGCTCAGTTGGTA
miR1142	382	AAGGTGTTGTGGTGCGGTCATGGG	miR2916	4881	GGGGCTCGAAGACGATTAGAT
miR1143	5	AGGACATCTCCCTTACGGA	miR2937	3	ATTAGATGCTGTTGAAGGAGT
miR1144	8	TCAGGCAGCGGCGGGATTCTGG	miR2939	2	TGACGCATCTAACACTAAGACACAT
miR1145	266	AGAGGCGTTGCACACCTGTCGGTGG	miR2948	1250	TAGTGCGAGTAGATGGGCAAGAATC
miR1146	177	ATAGGGTGACCGGAAGCT	miR2949	164	TCCTTTGATGGATTTGCCA
miR1147	6	TCATCGGCCAATGTCTGCGCGCAGA	miR2950	3	TGGTGTGCACGGGATGGAATA
miR1149	4	TACGGCAACTACACCACCCACAGC	miR319	471	GAGTTTTCTTCGGACATA

转下页

续表

miRNA 家族	表达量	序列(5'→3')	miRNA 家族	表达量	序列(5'→3')
miR1150	30	TGCTAGCAGCGACCTGGGGCCTA	miR3476	914	TGACTGGAGGATGTTGGCTGC
miR1151	7	ACGGTGTGGTGCCGGACCCGG	miR3508	79	TAGAGGTCCCCTGTTCTGA
miR1152	206	TCAGAAGCGGCTGTCTTGA	miR3511	836	CCGGGCAACGAATGCAGA
miR1153	34	AGATTGTACACGATGGCTC	miR3512	40	TGCAAATCATGACAATGA
miR1154	2198	CTTAGTCGATCCTAAGGCAT	miR3513	13	ATATTGATAGAAATTGTAT
miR1156	233	TTCAGACTGGCATCAGGCAC	miR3515	6	AATGGAAAAGTGAACGGTATA
miR1160	19	TCAAAGGACGCAGAGCGGAGT	miR3518	253	TCCTTTGGGTGATTTCTGTG
miR1161	141	TCTGTAGTTCATCAACAGT	miR3519	2398	TCAAACAATGACAGCATCTGATCCA
miR1162	404	CGGCTTAATTTGACTCAACACG	miR3521	374	TGTGTGATCGTATACATACATGT
miR1163	12	ACGTGCGCCATGGCACGCAGC	miR3522	1	TGAGACCAAATGAGCAGCTGA
miR1165	3000	TACCTACAAGCGGTCTGGAGCC	miR390	928	AAATAGTAGGGATAGCACC
miR1166	13	AGAGTCACGATGACCTCTGGG	miR395	3833	CTGAAGATCGGAGGACCAC
miR1167	99	GGGGCCTGATGATGAAAC	miR396	19	TTCCACAGCTTTCTTGAA
miR1168	2716	ACACGGACAAGGCGAAGAA	miR397	611	ATTTAGGTGCAGCGTCGTGA
miR1169	154	TTACGCCACGCATGCAACATCCACA	miR398	943	TGTGTTTCATCACGGACAGCACCTG
miR1171	32	TGTGGAGTGGAACGTGGAGTGG	miR399	155	TGCCAAAAGGAAATCGCTC
miR1172	31	AGGATTGCAGCAGCTGCAC- CGGGC	miR400	3	TATCGAGAGTATCTGATAAATACAC
miR1212	465	GTGGGAGACATAGAAGCG	miR403	4	TTAGATTCACGCACAAACT
miR1216	161	TGTGGTGATGCAGCTTGATT	miR408	3	CTGGGAACAAGGCAGATCCTGAA
miR1217	4321	TGGAACATGTTGGCAAATTGC	miR413	1	AAGTTTCTCTTGGTTCCGCC
miR1218	176	CCTTCAGGGTCGTAGCCTCAG	miR414	1580	TATCTCATCGTCATCGCCA
miR1221	1	TGGTGTGTGCCAGGTCAAC	miR415	31	CACAGAACTGAAGAAGCAG
miR1222	3648	TTGAAAGGGACTTGCAGTGGTACAC	miR416	338	TGTGGGTCCGTCACTGTTC
miR1313	17	TTCCCTGAATTATTGTTGCG	miR418	188	TGAGTGATGATGAAATGACAGC
miR1318	647	TCAGGAGATGATGACATAGAAC	miR419	1049	TATGACTGATGACGTGTT
miR1424	35	ATGCAGATGATGCTGATCGT	miR426	225	TTTTGGACATTGTGCTGAC
miR1426	16	AGAAGTTATGATGATTAAG	miR435	89	TTTATCGGTATTGGATGA
miR1427	9	TGCGGGAACCGTGCGGGTGGC	miR437	917	AAAGGTAGAAAGTTTGCTT
miR1428	3	AAGATGAATGCACATGAATGTTGA	miR440	7	AGGGTCTCCTCGATGATGCTGGACA
miR1431	11	TTTGCGTGTTGGCCGTCGC	miR443	837	ATACATGCAAAAAATCTGGA
miR1432	137	ATCAGGAGGAACACCGACA	miR444	212	TGCAGTTGTTGCTGCCCAAGTT
miR1441	9341	ACCGAAGTGGAGAAAGGTTT	miR474	147	CATAAGTTTTTGGGTTCTGGGGG
miR1444	82	TCCACCTCGGGCAATGTT	miR477	9	TCCCGTCAAGGCTTTCCA
miR1450	133	TTCAATGATGCTGCGGTCGGTTTC	miR479	3	TGTGGTATTGGTTCGGCTCATC
miR1507	445	ATGTGTTGTATGGAACGAACACAGT	miR482	163	TGCGGAACATGAAGGAGGCTT
miR1509	3	TTAATCAGGGAAATCACAGTT	miR528	426	TTGAGGGGATCAGAGGAG
miR1510	11	TGGAGGATCAGGTAAACAAC	miR529	175	AGGAGAGAGAGCACTGCC
miR1513	6	ATGAGGACAGCCAGACTTAC	miR533	1784	CTGACACAGGTGCTGCATGGCTGTC
miR1514	2	TTATTTTGAAAATGAAGGCTATTGA	miR535	8	TGGACAGACGAGAGAGTGCAGCGGG
miR1516	663	CAAAAGAGCGTATGGGTA	miR536	5	TGTGTCAAGCTGTGTGGCGAC
miR1520	15489	TCAGAACTGAAACGGACAA	miR538	9	TTGCCTGCGAGTCTAAGTCTGAGAA

转下页

续表

miRNA 家族	表达量	序列(5'→3')	miRNA 家族	表达量	序列(5'→3')
miR1528	31	ATAGGATTTAGGTCAATATATTT	miR775	197	TCGATGTGAGTAGTGCCA
miR1531	4819	ACGTCCACTCTGGGAAGACGTTGT	miR780	24	TCAAGCAGCTGTATGAACAGGCT
miR1533	16	TATAATAAAAGATAAAAGGA	miR783	4	ATGCTGGCCGGTTCATTTT
miR1535	139	TCTTGTTTGTGATAGCATGTCT	miR807	1	TCGGCATTCCCAGGTGAATCC
miR156	489	TGACAGAAGAGAGTGAGCAC	miR810	80	AAGTGATTTAGGTATGCGT
miR157	493	TTGACGGAAGATAGAGAGCAC	miR815	6	AAGGGGAGAGGAGAATTGG
miR158	129	TCCCAGAACTTGTACAGACAAAGCA	miR816	194	GTGCACCTTTACTACAAC
miR159	586	ATTGGATTGAGAGGGCAGCCA	miR822	312	TGCGAAAGCATTTGCCAAG
miR160	187	TCATAAGAGGTCACGCAGG	miR825	50	TTCTCCGAAGGTGCATGACC
miR161	2	TCGAAAGTGAAGTGAAGATGACTA	miR827	3977	TTAGAGGACGATCAGCAAC
miR162	1	TCGATAAACCTCTGCATCCAG	miR829	1	ATGCTCTGATACCAACTGATGGAC
miR164	4	TGGAGAAGCAGGGCACGTGCA	miR832	64	TTTGGGATCGGGAATAAA
miR165	16	TCGGACCAGGCTTCATCCCC	miR833	3800	TGTTTTTGACTCGGTATAGA
miR166	220	TCGGACCAGGCTTCATTCCCC	miR834	484	GTGGTAGCAGAGCGCTGTAA
miR167	155	ATGAAGCTGCACTGCCATTATCTGG	miR835	12	TTGAGAAGATCGTAGAAAG
miR168	43	TCGCTTGGTGCAGGTCGGGAA	miR838	48	TTTTCTTTTCTTCTTGACAA
miR169	1294	TAGCCAAGATATTGACGTTGGCCTG	miR843	1575	TTTAGGTAGAGCCTCATGA
miR171	106	TGGATCTCTCGGCTCATGT	miR846	2	CTGCAATTGAAGGCTATGAATT
miR172	329	TGAGATCTTGAAGGCATGCTGCAGC	miR850	2	CTAAGATCTGGCTACGAACAACAG
miR1846	181	TATCCGGCGTGCATGGGG	miR854	22	ATGAGGAGAGGGGGGAGGG
miR1853	3082	TATTGGGCATGATTTCGGTTGACAT	miR857	14	TTTTGTATGTAGGAGGTTA
miR1854	235	GGTGAAATTGTAGATCGG	miR860	3	TCAATAGATTGGACTTCTTTATC
miR1855	60	AGCAATGGAGTAGCCACGAGTATGA	miR863	5	TGGAGTAGAACAAGACATTAT
miR1856	10	TAACGGAGACGGATTCGTA	miR864	4	TCAGGTTGATCTGAACTTCAAGAA
miR1857	289	TCAGGCGTACAAGAAAACCTCAGG	miR865	11	ATAAATTTGGATAAATTGA
miR1858	2684	TAGAGGAGGACGGAGTGAGTAGGCT	miR866	188	ACAAAATCGTCATCGAAG
miR1861	65	TGGTCATGAAGAGCAGGAACCTTGA	miR869	28	TCTGGTGTGTGTGTGTTGC
miR1874	11140	TAGGAAGTAGGTGTAACCCGCATG	miR894	10	TTTCACGTCGGGTTACCA
miR1875	2637	ACAGATGGAATGACTGCAACAGAAG	miR896	112	GTCAGTTTGGCCGAGTGGTTAAGG
miR1877	3962	AATGACATGTGTAGGATAGGGG	miR900	6	TCCCAGGATAAAGAACCCGGC
miR1878	6	TACCTAATCTGATACTATAAAAAG	miR902	232	AGAGGATCTCAACATTGTAC
miR1881	472	ACATGTTAGTTGTAGCTGGTGTAT	miR904	1453	TCTTGTCACATGGAATGTAGGGGCA
miR1884	1211	TGTGACGGCGTGCCTTTTGCAT	miR906	8	CGAGTTGGTCGGGCGTGAAGC
miR1886	38	TGGAGATGACATCTTCGATTGGG	miR908	1	ACTACGTCAACAGAGTCAG
miR1917	1	TTACCAAAGAGTCTAAAGT	miR910	93	ACACCTCGGGCTCGACGC
miR1918	211	TGTAGGTCGTGAGTTCGATTCTC	miR911	421	ACAATGGTGTCCGGACGATTT
miR1919	822	ACGAGAGTATTTGGAAGG	miR913	6	TGCCACTTCAAGTCCGTGAG
miR2079	290	AGAAGATTGATGTTGAGAGACTGCA	miR916	2993	CAAGAGGTCATCGGTTTCGAC
miR2083	30	ATATTGGAGTGCAGAGAAGA	miR917	69	TTCAAGGGATGTTTCAAG
miR2086	3	GACATGAATGCAGAACTGGAA	miR919	114	TCTAGGAGGACGCTCACCCT
miR2088	1	TCCAATGTAATCTAGGTCTAC	miR946	9	TGTGGATAGAGATGGTGATGT
miR2092	172	CAACTGATCGGGTTCCT	miR947	9469	ACATCGGAATGTGTATACTAGTGTC
miR2095	19	ATGATAATCTTTACGTGAATA	miR948	52	TCTAGGCTGTCTGTTGGATTTCGG
miR2097	193	AGAGACTGAGACGGAAGGGAAG	miR949	176	TGACTCCGGAAGCCGAATGCTGCC
miR2098	83	CGGTTTGTTAAGTCGGATG	miR951	2	TTGCTTGACGGTCTGGACGAC
miR2099	27	TGAACAAGTTTGGTACAAGTTT	miR952	2850	AACGAGGATCCATTGGAG
miR2108	125	TTAAAGTGTTGTAGTGTGCGG			