**生物序列标准文件抽检**

1. **基本信息**

抽检数据：FASTA-EP中xml文件（共27个）

抽检时间：2015年6月12日-6月16日

1. **抽检问题**

抽检主要发现了三种问题：

1. 专利下的序列数量不全
2. 具体序列Seq id中的md5值与网络检索值不一致。
3. FASTA-EP和网站的信息一致，但是比专利原文中多出一条序列，可能为网站错误。
4. **具体结果**

抽检量：3件xml文件中的6条专利

1、1b2ad6e5-9b1b-46da-a753-52e3d41dbc52.xml

（1）PNO: EP1202065A1

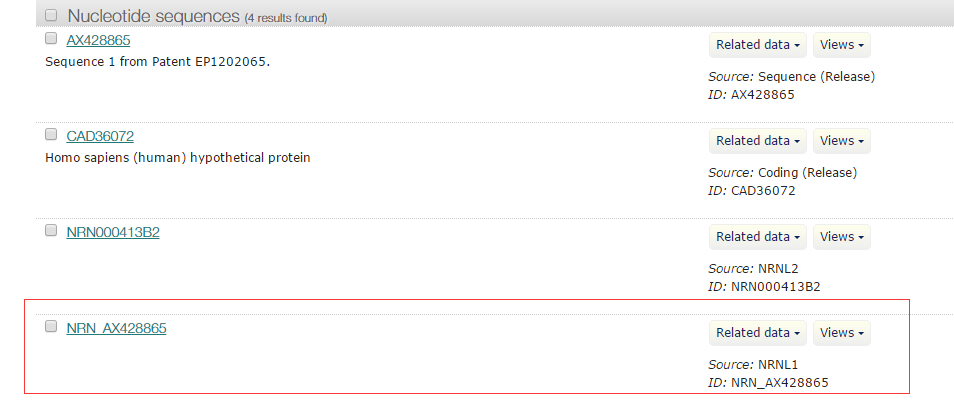
|  |  |
| --- | --- |
| **序号** | **问题描述** |
| 1 | 没有序列AX428876 |
| 2 | AX428866 seq id中的md5与embl文件中md5值不一致 |
| 3 | AX428868 seq id中的md5与embl文件中md5值不一致 |
| 4 | AX428865 seq id中的md5与embl文件中md5值不一致 |
| 5 | AX428867 seq id中的md5与embl文件中md5值不一致 |

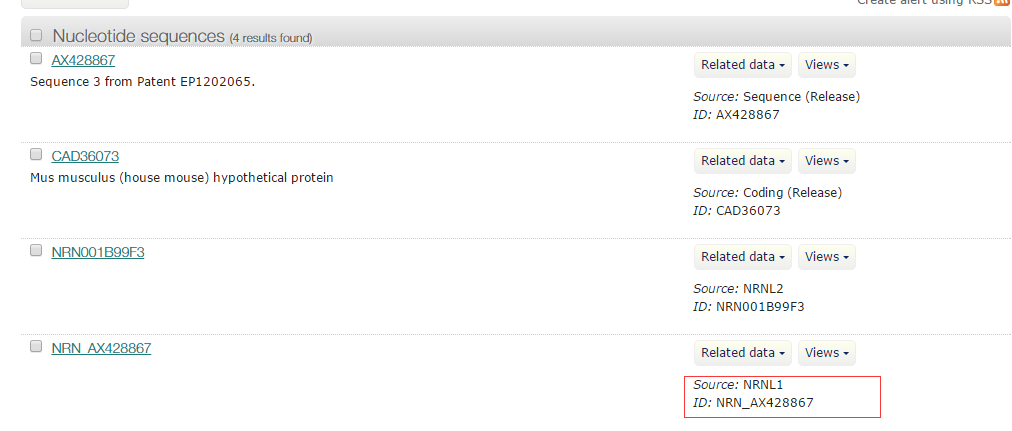
*Jiahh:Source:*Sequence (Release)





*Jiahh:*AX428865 其实没有，是根据NRN\_AX428865截取后匹配上的；867同理

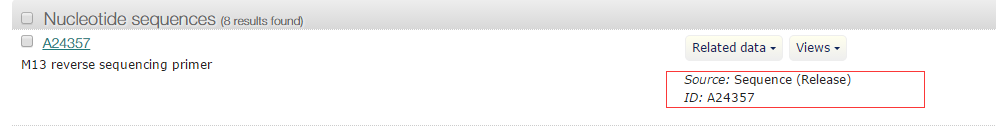




3（2）PNO: EP0524768A2

|  |  |
| --- | --- |
| **序号** | **问题描述** |
| 6 | A24412 seq id中的md5与embl文件中md5值不一致 |
| 7 | 没有序列A24357，该序列为专利原文中NO.5序列 |
| 8 | 原网站中提供了A24754/NO.43条序列;在专利原文中没有找到第43条序列号；可能为网站错误 |

*Jiahh:Source:*Sequence (Release)



2、9430a94c-0db1-45fe-a6e8-349af46a8522

（1）PNO: EP2297186A2

|  |  |
| --- | --- |
| **序号** | **问题描述** |
| 9 | 没有序列JA398200和JA398221； |
| 10 | JA398213 seq id中的md5与embl文件中md5值不一致 |
| 11 | JA398225 seq id中的md5与embl文件中md5值不一致 |
| 12 | JA398230 seq id中的md5与embl文件中md5值不一致 |
| 13 | JA398210 seq id中的md5与embl文件中md5值不一致 |
| 14 | JA398205 seq id中的md5与embl文件中md5值不一致 |
| 15 | JA398211 seq id中的md5与embl文件中md5值不一致 |
| 16 | JA398219 seq id中的md5与embl文件中md5值不一致 |
| 17 | JA398229 seq id中的md5与embl文件中md5值不一致 |
| 18 | JA398222 seq id中的md5与embl文件中md5值不一致 |
| 19 | JA398220 seq id中的md5与embl文件中md5值不一致 |
| 20 | JA398231 seq id中的md5与embl文件中md5值不一致 |

*Jiahh:Source:*Sequence (Release)



*Jiahh:格式2数据，导出本xml时，格式2未导入*



（2）EP2406730A1

|  |  |
| --- | --- |
| **序号** | **问题描述** |
| 21 | 专利生物序列表共20条序列，网页检索到的公开序列有20，缺少序列ID为：JA735755、JA735757、JA735759、JA735761、JA735763、JA735764、JA735765、JA735766、JA735768、JA735770、JA735771、JA735772、JA735773、JA735774 |
| 22 | JA735756 seq id中的md5与embl文件中md5值不一致 |

*Jiahh:*均为*Source:*Sequence (Release)

3、f77d3cbf-3369-4571-8aba-6c1d8d39d01d

（1）PNO: EP2245148A1

|  |  |
| --- | --- |
| **序号** | **问题描述** |
| 23 | 专利序列表中共161条序列，本文件中96条，缺失65条序列 |
| 24 | 抽检：HI644862 seq id中的md5与embl文件中md5值不一致 |

（2）PNO: EP2413968A1

|  |  |
| --- | --- |
| **序号** | **问题描述** |
| 25 | 专利中共119条序列，本文件中117条，缺失2条序列 |