**国外生物序列数据抽检记录**

**一、抽检文件名**：ipphdata20150706

**二、抽检方式**：抽6个xml，每个xml抽两个专利号，每个专利号抽2-3个生物序列进行质检。

**三、抽检发现问题**：

1. 网站本身问题，EMBL1非标准文件，没有提供MD5值，所以抽检中这些文件的MD5值未进行比对。
2. 网站本身问题，EMBL1非标准文件，描述序列的特征，仅提供的<location>和<keywords>，没有<other>字段。
3. 部分EMBL1非标准文件，缺失<organism>，需要进行提取；

提取方式：取标识符“OS”后面的所有值，去掉首尾空格。

**四、具体抽检记录**：

|  |  |  |  |
| --- | --- | --- | --- |
| **文件** | **抽检专利号PNO** | **抽检序列ID** | **问题描述** |
| 1436160764433-0 | FR2599752 |  | 无 |
| DE19846493 |  | 无 |
| 1436161647896-0 | US7425612 |  | 无 |
| US7511190 |  | 无 |
| 1436162540147-0 | JP2008513004 | DD756530 | 1、源文件没有提供MD5值，未进行核对 2、无organism字段 |
| JP2008212152 | DD815756 | 无organism字段 |
| 1436164191192-0 | JP2010268796 | DJ586543 | 无organism字段 |
| JP2000511771 | DD919394 | 无organism字段 |
| 1436165295731-0 | JP2014110793 |  | 无 |
| 1436166322837-0 | US7122328 | ABL15877 | 无organism字段 |
| US5976877 | AAE34079 | 无organism字段 |
| AAE34078 | 无organism字段 |