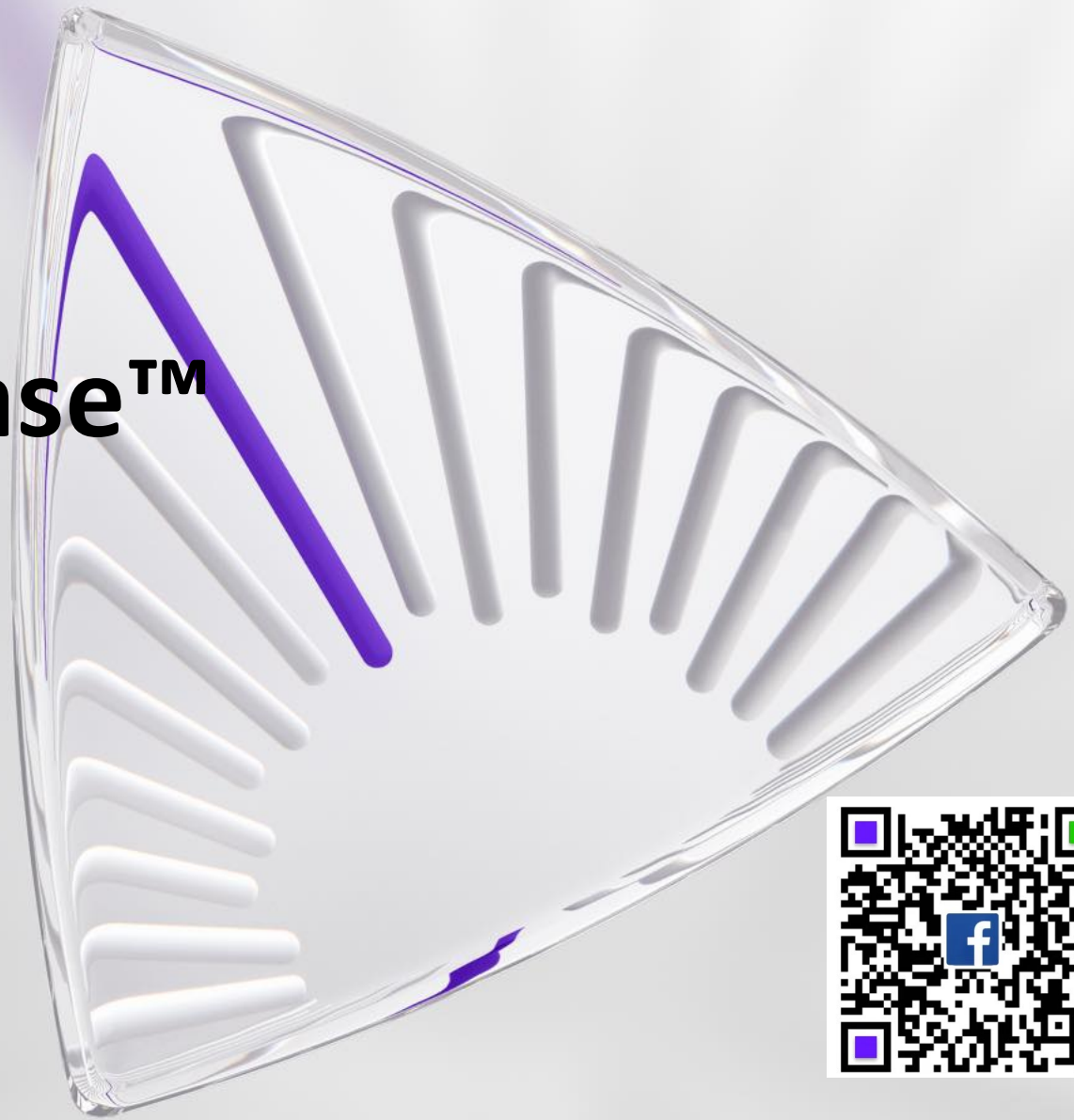
Clarivate
Derwent™

# Derwent SequenceBase™
# 使用者快速上手指南

**2020年7月版本**



歡迎關注我們的粉絲團

https://clarivate.com.tw/

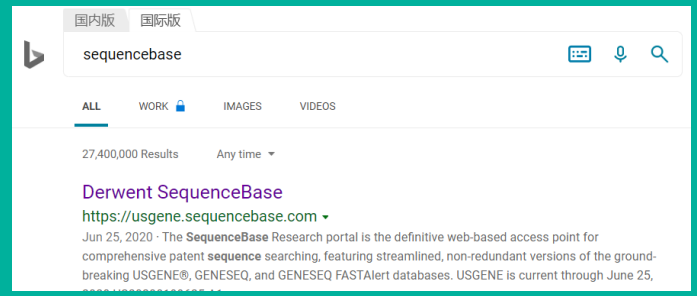
# 如何登入？

請通過如下網址進入Derwent SequenceBase產品登入介面



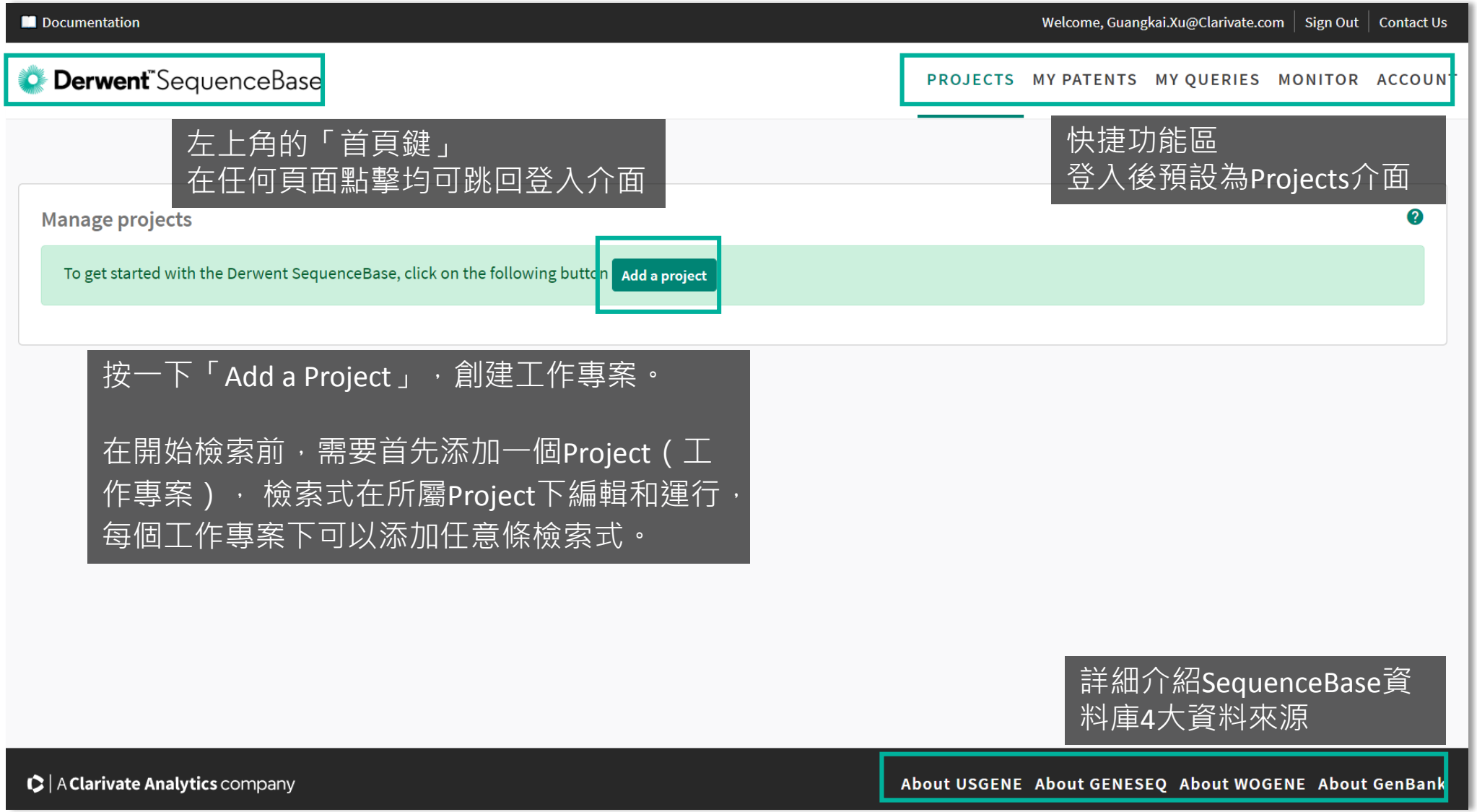

## 附註說明

Derwent SequenceBase是在雲端運行的資料庫，無需安裝。使用者可將產品登入頁面添加為「書籤」。或通過搜尋引擎搜關鍵字SequenceBase

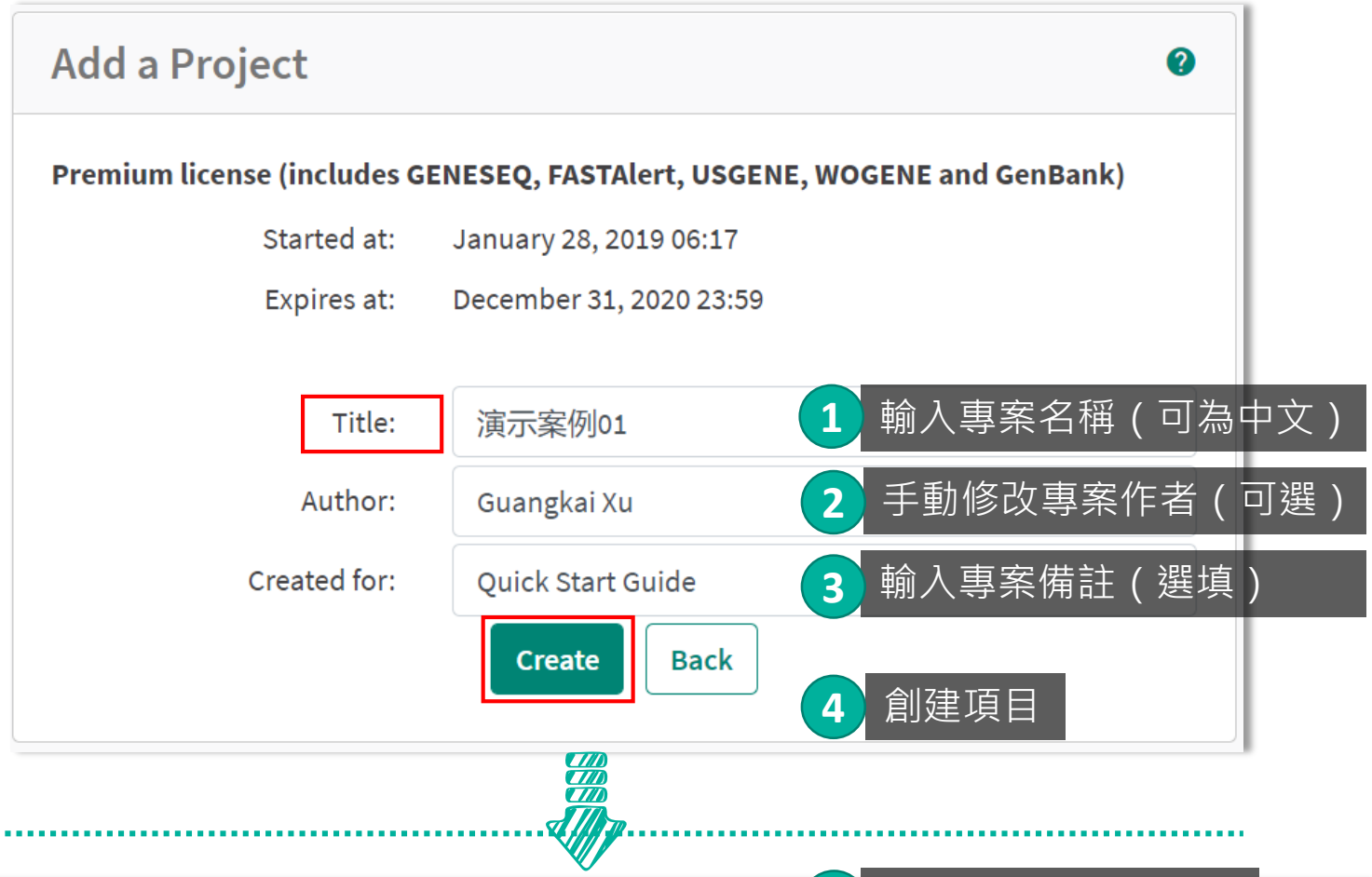


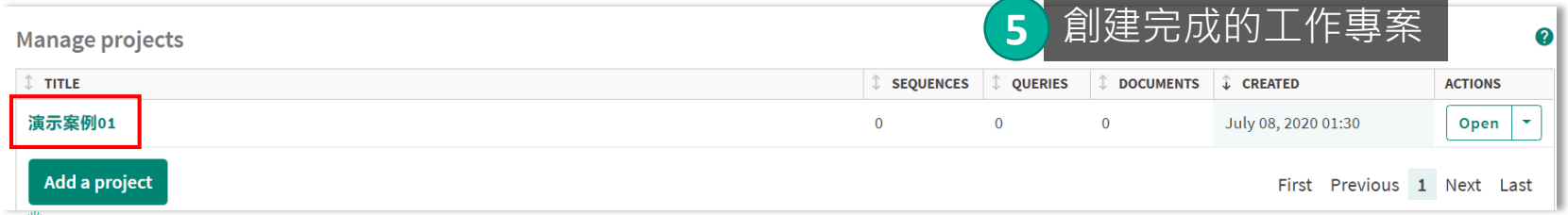

Derwent SequenceBase一個帳號對應一個用戶。同一帳號，後登入者會將前面的登入者踢出。

# 登入介面


Documentation
Welcome, Guangkai.Xu@Clarivate.com
Sign Out
Contact Us
Derwent SequenceBase
PROJECTS
MY PATENTS
MY QUERIES
MONITOR
ACCOUNT
左上角的「首頁鍵」
在任何頁面點擊均可跳回登入介面
快捷功能區
登入後預設為Projects介面
Manage projects
To get started with the Derwent SequenceBase, click on the following button
Add a project
按一下「Add a Project」，創建工作專案。
在開始檢索前，需要首先添加一個Project（工作專案），檢索式在所屬Project下編輯和運行，每個工作專案下可以添加任意條檢索式。
詳細介紹SequenceBase資料庫4大資料來源
A Clarivate Analytics company
About USGENE
About GENESEQ
About WOGENE
About GenBank

# 創建工作專案




## 附註說明

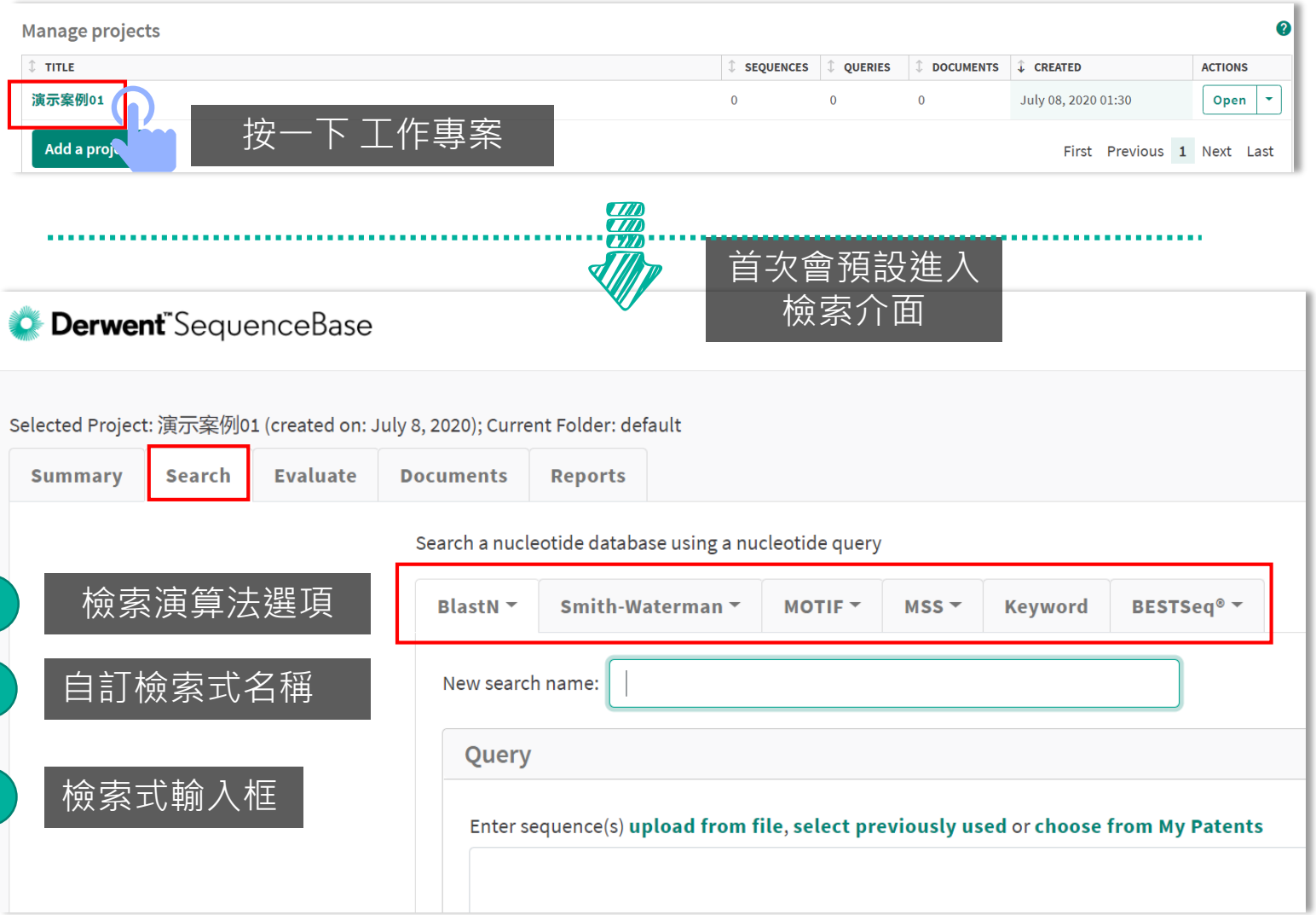
[Project] 工作專案可以理解為檢索的集合。添加一個專案來執行序列檢索、瀏覽結果和生成報告。

每個項目下可以運行無限條檢索式，匯出上限為10000個文檔。

請注意，在兩種情況下，專案將自動關閉：

- 超出了允許的檔數量——專案檔案夾中的檔數量超過9999個。
- 您的帳戶許可證已過期。

# 工作專案——檢索




## 附註說明

在檢索介面，檢索需要嚴格按照以下步驟進行：

1、選擇適當的檢索演算法；

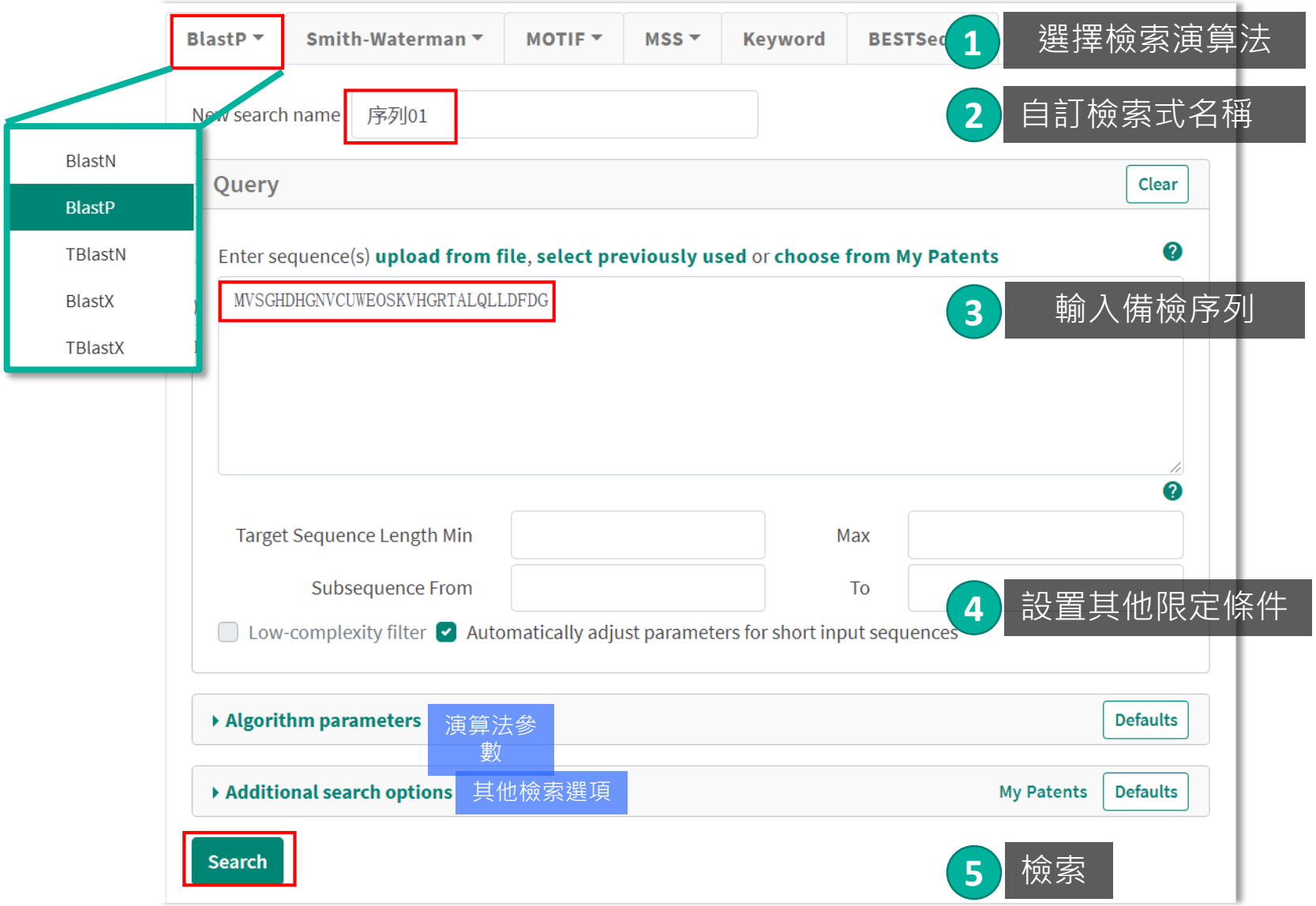
2、輸入自訂的檢索式名稱；

3、在檢索式輸入框內輸入FASTA格式的序列；

4、設置檢索的其他限定條件；

5、檢索。

# 工作專案——檢索演算法（Blast）




## 附註說明

SequenceBase支持6種檢索演算法

其中Blast，Smith-Waterman，MOTIF，MSS和BESTSeq這5種檢索演算法可用於檢索核苷酸和蛋白質序列。

**Basic Local Alignment Search Tool（Blast）是NCBI進行序檢索的演算法**，其優點為檢索速度快，結果比較精準。

# 工作專案——檢索演算法（Blast）

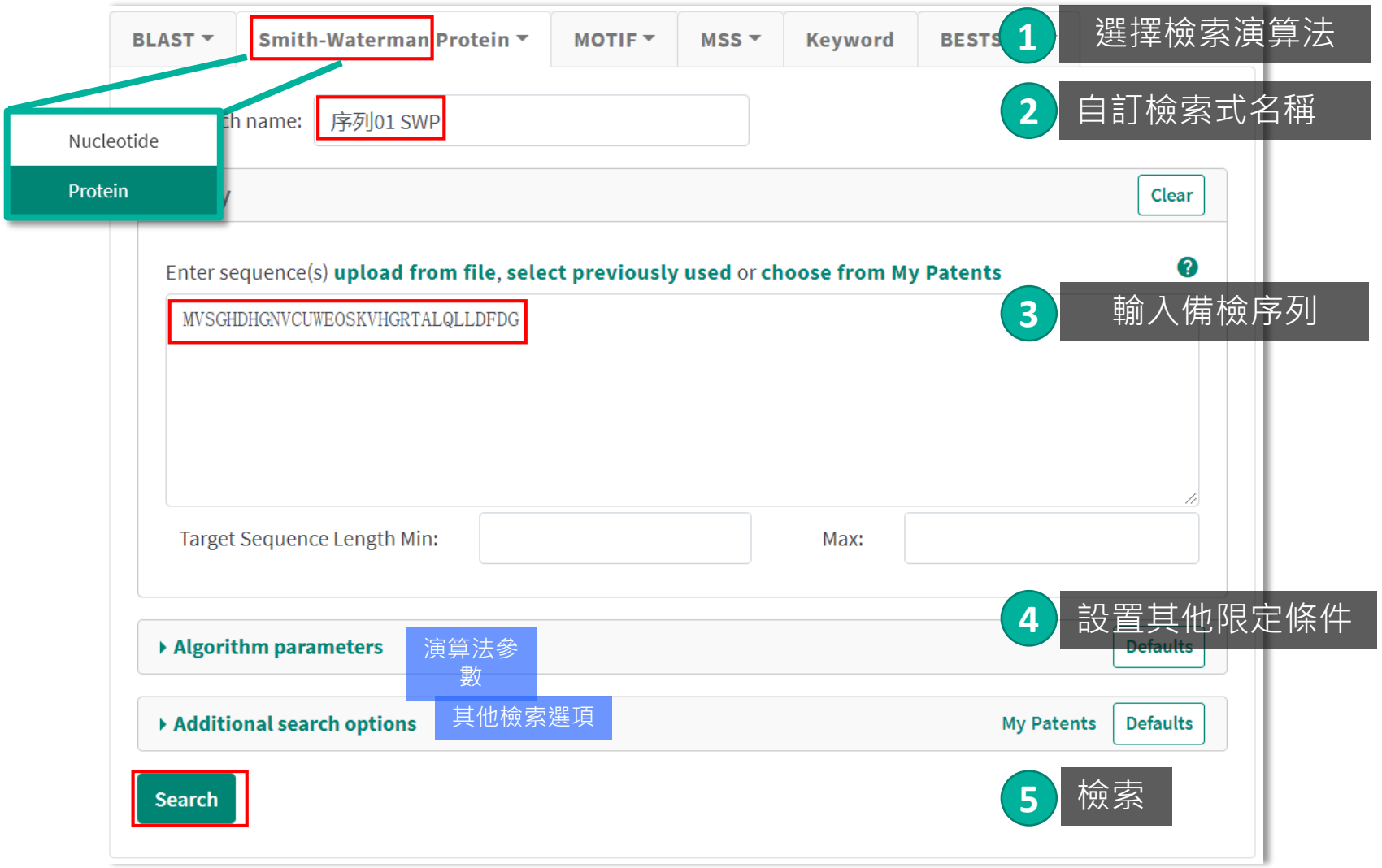
| | |
|---|---|
| BlastN | 將一個DNA查詢序列不同方向的兩條鏈與DNA資料庫比較 |
| BlastP | 將一個蛋白質查詢序列與蛋白質數據庫比較 |
| TBlastN | 使用蛋白質序列在翻譯的核苷酸數據庫中檢索<br>將一個DNA資料庫中的每一條序列翻譯成6種可能的蛋白質，然後將要查詢的蛋白質序列與翻譯的蛋白質逐一進行比較。 |
| BlastX | 使用翻譯的核苷酸序列在蛋白質數據庫中檢索<br>將一個DNA序列用所有可能的閱讀框翻譯成6個蛋白質，然後將它們逐一與一個蛋白質資料庫進行比較。 |
| TBlastX | 使用翻譯的核苷酸序列在翻譯的核苷酸數據庫中檢索<br>將查詢DNA及資料庫中的DNA都翻譯成6種可能的蛋白質，然後進行36次蛋白質-蛋白質資料庫檢索。 |

## 附註說明

**Blast演算法**包含5種細分演算法：

- BlastN
- BlastP
- TBlastN
- BlastX
- TBlastX

# 工作專案——檢索演算法（Smith-Waterman）

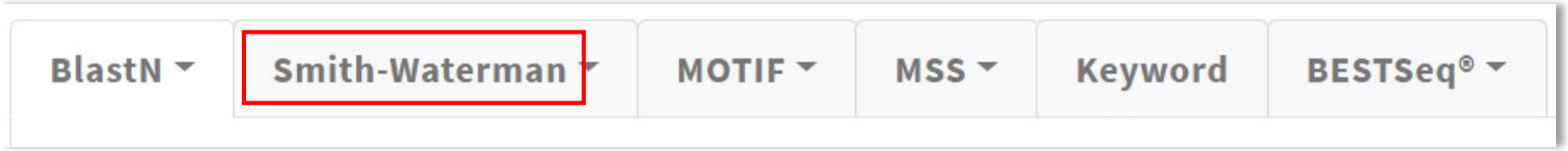



## 附註說明

SequenceBase支持6種檢索演算法其中Blast，Smith-Waterman，MOTIF，MSS和BESTSeq這5種檢索演算法可用於檢索核苷酸和蛋白質序列。

Smith-Waterman是常規免費序列檢索平臺不提供的檢索演算法，包含核苷酸和蛋白質兩種細分。

# 工作專案——檢索演算法（Smith-Waterman）




Nucleotide　使用核苷酸序列在核苷酸資料庫中檢索

Protein　使用蛋白質序列在蛋白質數據庫中檢索

Smith-Waterman演算法是一種用來尋找並比較具有局部相似性區域的動態規劃演算法，後來很多演算法都是在該演算法的基礎上發展起來的。

這是一種兩序列局部比對演算法，把兩條未知的序列進行排列，允許刪除和插入操作。這種比對方法可能會揭示一些匹配的序列段，而本來這些序列段是被一些完全不相關的殘基所淹沒的。

## 附註說明

SequenceBase支持6種檢索演算法其中Blast，Smith-Waterman，MOTIF，MSS和BESTSeq這5種檢索演算法可用於檢索核苷酸和蛋白質序列。

Smith-Waterman是常規免費序列檢索平臺不提供的檢索演算法，包含核苷酸和蛋白質兩種細分。

# BLAST

# vs

# Smith-Waterman

**BLAST**是一種近似演算法，因其**速度和效率**而得到普及。

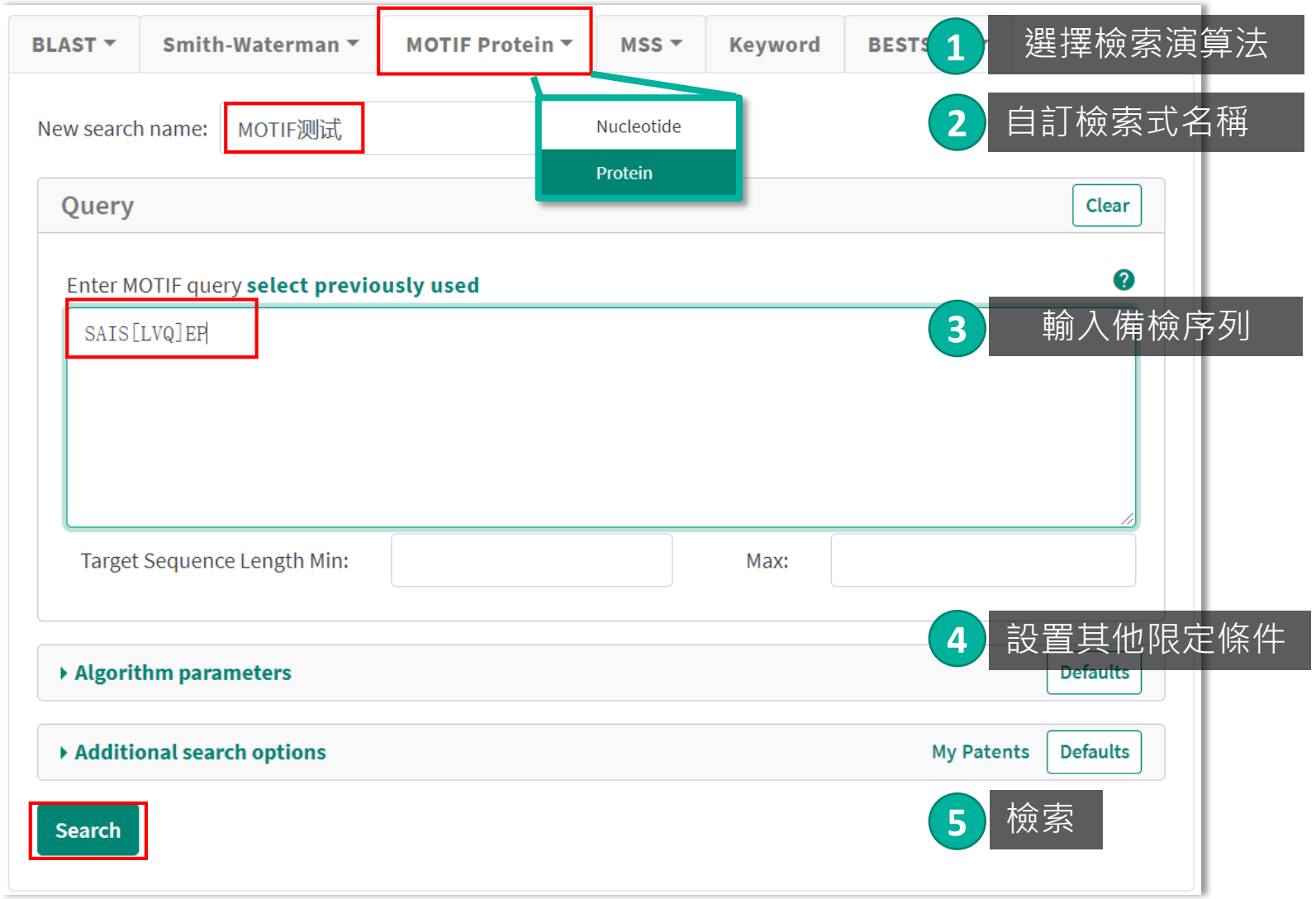
**Smith-Waterman（SW）**屬於動態規劃演算法中的局部動態規劃演算法，該演算法是**最精確的一種序列比對演算法**[1]。用於發現兩條序列在局部水準上的相似性。當研究人員更關注檢索結果匹配精準程度的時候，更推薦使用SW演算法。

如果您希望知道："我們應該使用Smith-Waterman還是應該使用BLAST？"
**我們建議同時使用兩種演算法檢索，以提升檢索結果的全面性。**

傳統的SW演算法的缺點是某些長序列的檢索需要花費大量時間。但SequenceBase開發的專門的基因比對搜索技術，以及雲計算的進步，讓我們在SequenceBase平臺運行SW查詢所需的時間減少到與BLAST相當。

*[1] 陳銘.生物資訊學[M].第三版.北京:科學出版社,2018:50.*

# 工作專案——檢索演算法（MOTIF）




## 附註說明

MOTIF，可變功能域檢索，包含核苷酸和蛋白質兩種細分。

SequenceBase獨具特色的檢索功能。

可以實現指定殘基取代、指定位點取代、殘基重複n次出現以及序列片段組合等特殊檢索需求。

具體檢索指定參見下頁。

# 工作專案——檢索演算法（MOTIF）

| 檢索符 | 檢索目的 | 檢索式示例 | 檢索結果 |
|---|---|---|---|
| [] | 指定取代的殘基 | SAIS[LVQ]EP | SAISQEP<br>SAISVEP<br>SAISLEP |
| [-] or [~] | 排除一個或多個殘基 | SAIS[-Q]EP | SAISVEP<br>SAISEEP |
| ^ | 在序列的啟點或末尾處檢索目標序列 | ^ATGCGK<br>DATGC^<br>^ATGCGK^ | ATGCGKTTI<br>EDELDATGC<br>ATGCGK |
| . | 序列中指定位點的替換 | RPVQ.RI | RPVQARI<br>RPVQGRI |
| \| | 指定取代的殘基 | SSLS\|NAIS<br>A(SSLS\|NAIS)G | NAIS<br>ASSLSG |

## 附註說明

MOTIF，可變功能域檢索，包含核苷酸和蛋白質兩種細分。

SequenceBase獨具特色的檢索功能。

可以實現指定殘基取代、指定位點取代、殘基重複n次出現以及序列片段組合等特殊檢索需求。

具體檢索指定參見左側。

# 工作專案——檢索演算法（MOTIF）

| 檢索符 | 檢索目的 | 檢索式示例 | 檢索結果 |
|---|---|---|---|
| ? | 目標殘基重複出現0次或1次 | NSK(VV)? | NSK<br>NSKVV |
| + | 目標殘基重複出現1次或多次 | NSK(VV)+ | NSKVV<br>NSKVVVV |
| * | 目標殘基重複出現0次或多次 | NSK(VV)* | NSK<br>NSKVV<br>NSKVVVV |
| {x} | 目標殘基重複出現 x 次 | AT(GC){2} | ATGCGC |
| {x1,x2,x3} | 目標殘基重複出現 x1 或 x2 或 x3 次 | AT(GC){1,3,4} | ATGCGCGC |
| {x1-x2} | Repeat from x1 up to x2 times | AT(GC){3-5} | ATGCGCGCGC |
| & | 將多個序列片段連接在一個序列中 | RSVK&KPLS&KFK | RSVK...KPLS...KFK |

## 附註說明

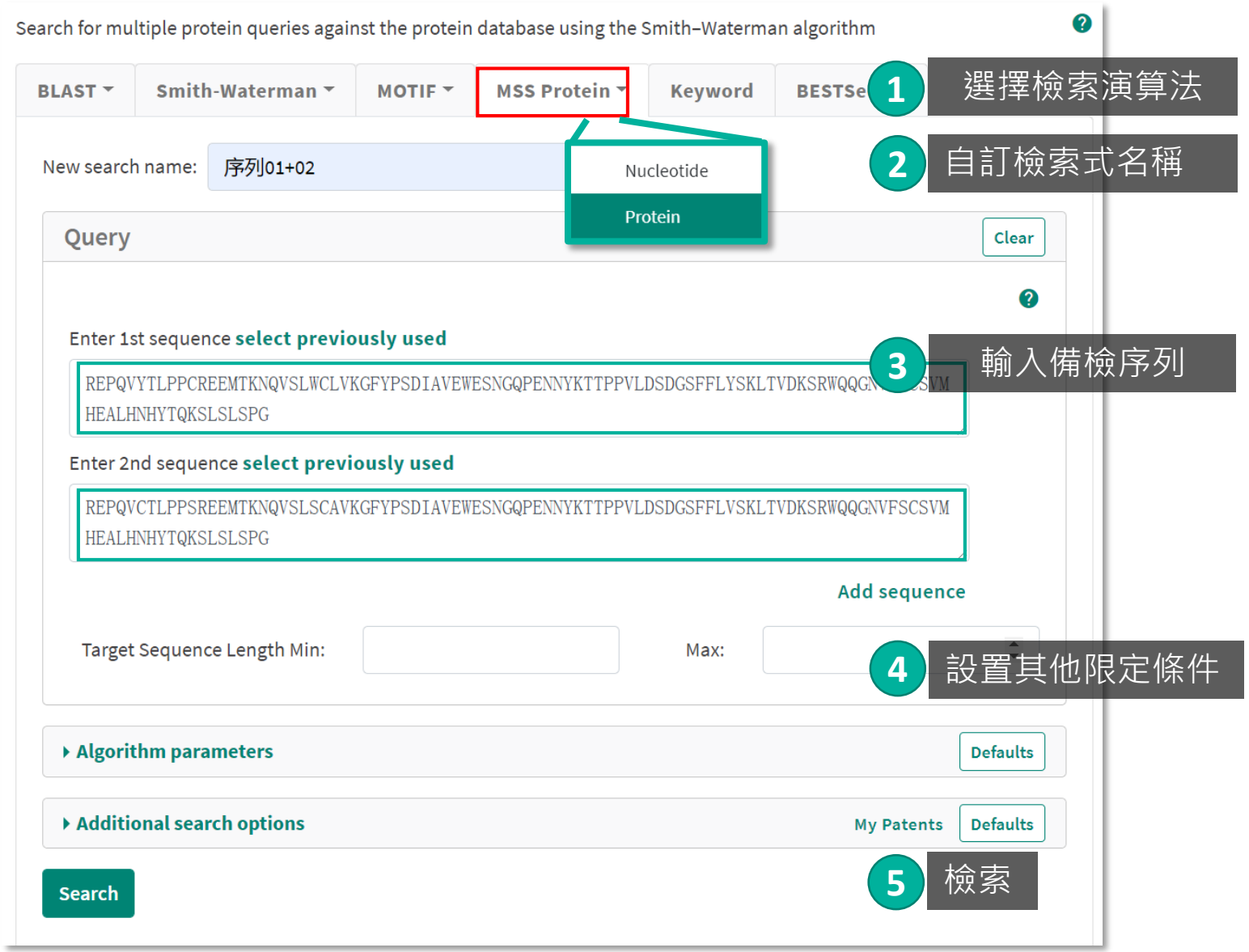
MOTIF，可變功能域檢索，包含核苷酸和蛋白質兩種細分。

SequenceBase獨具特色的檢索功能。

可以實現指定殘基取代、指定位點取代、殘基重複n次出現以及序列片段組合等特殊檢索需求。

具體檢索指定參見左側。

# 工作專案——檢索演算法（MSS）




## 附註說明

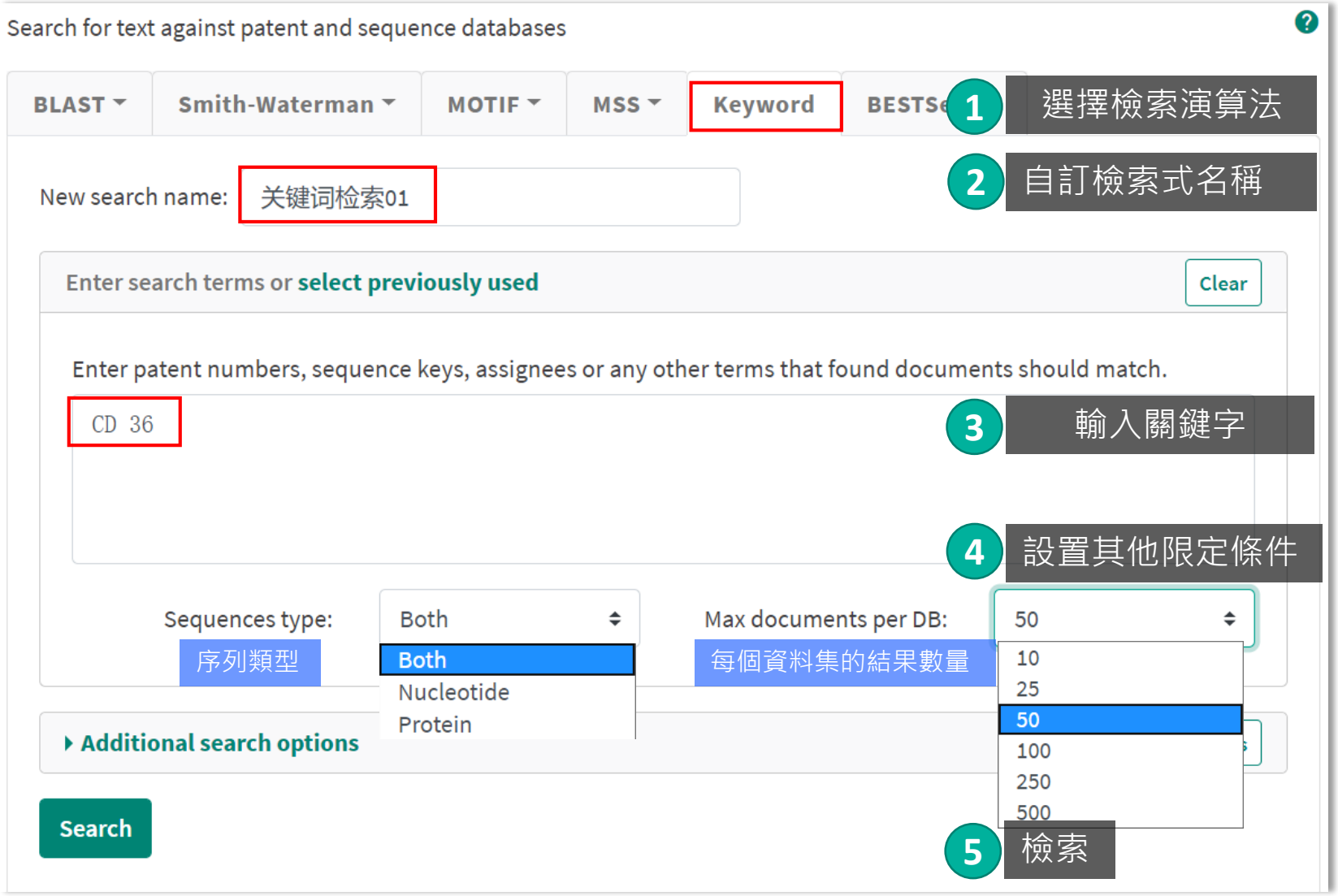
Multiple Sequence Search（MSS），多序列檢索功能。

SequenceBase獨具特色的檢索功能。

可以同時檢索最多6條序列，並得到同時存在多條被檢索序列的專利檔。

MSS中序列的檢索演算法採用Smith-Waterman演算法。

# 工作專案——檢索演算法（關鍵字）




## 附註說明

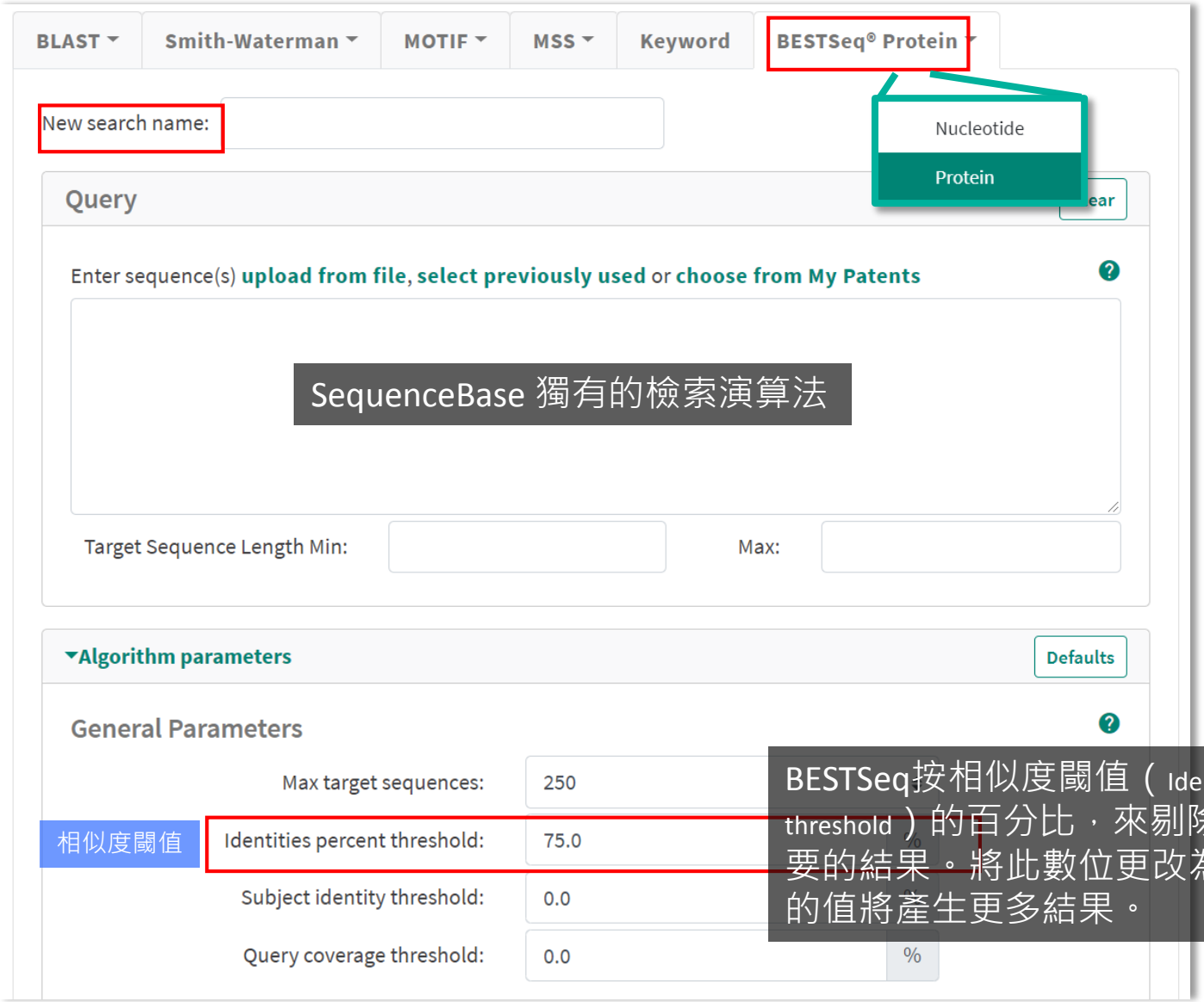
關鍵字檢索

您可以通過技術術語（如適應症、靶點等）、專利號、序號或專利權人等關鍵字來檢索專利和序列。

輸入關鍵字後可以進一步選擇被檢索序列的類型以及每個資料集的檢索結果數量。

（所述”每個資料集”分別為：GENESEQ/USGENE/WOGENE，例如當選擇50時，每個資料集最多可檢索得到50個文件）

# 工作專案——檢索演算法（BESTSeq）




BESTSeq按相似度閾值（Identity threshold）的百分比，來剔除不重要的結果。將此數位更改為較小的值將產生更多結果。

## 附註說明

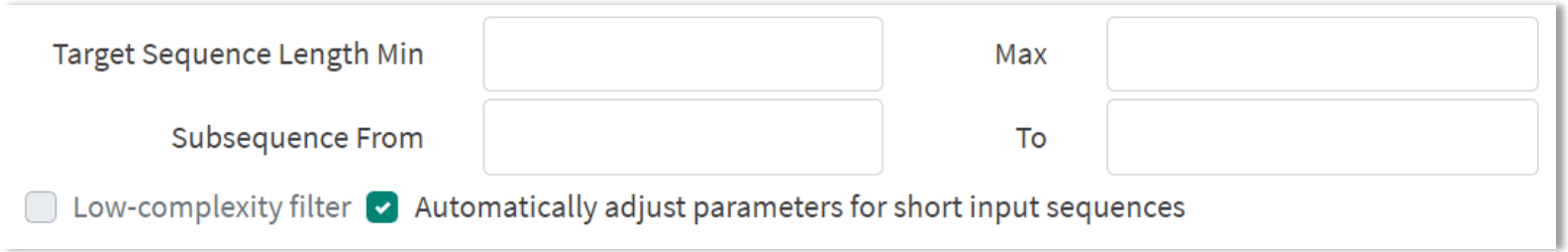
BESTSeq是SequenceBase獨家開發的序列檢索演算法。

目的是發現具有精確序列相似性的序列。它是可專利性檢索、專利無效檢索和FTO檢索的理想選擇。

BESTSeq依賴於序列間的替換、刪除、添加和/或插入數量的差異來檢索。它與Blast或Smith-Waterman演算法完全不同，即不應用任何估值矩陣，或類似的不相關參數。

# 工作專案——檢索演算法參數

**Blast演算法通用參數**




**Automatically adjust parameters for short input sequences**

自動調整短序列的輸入參數

勾選後將自動調整字長和其他參數，以改善Blast演算法檢索短序列的檢索結果。

**Low-complexity filter**

低複雜度篩檢程式

勾選後，可以過濾掉Blast演算法檢索結果中具有統計意義，但沒有生物學意義的結果。

## 附註說明

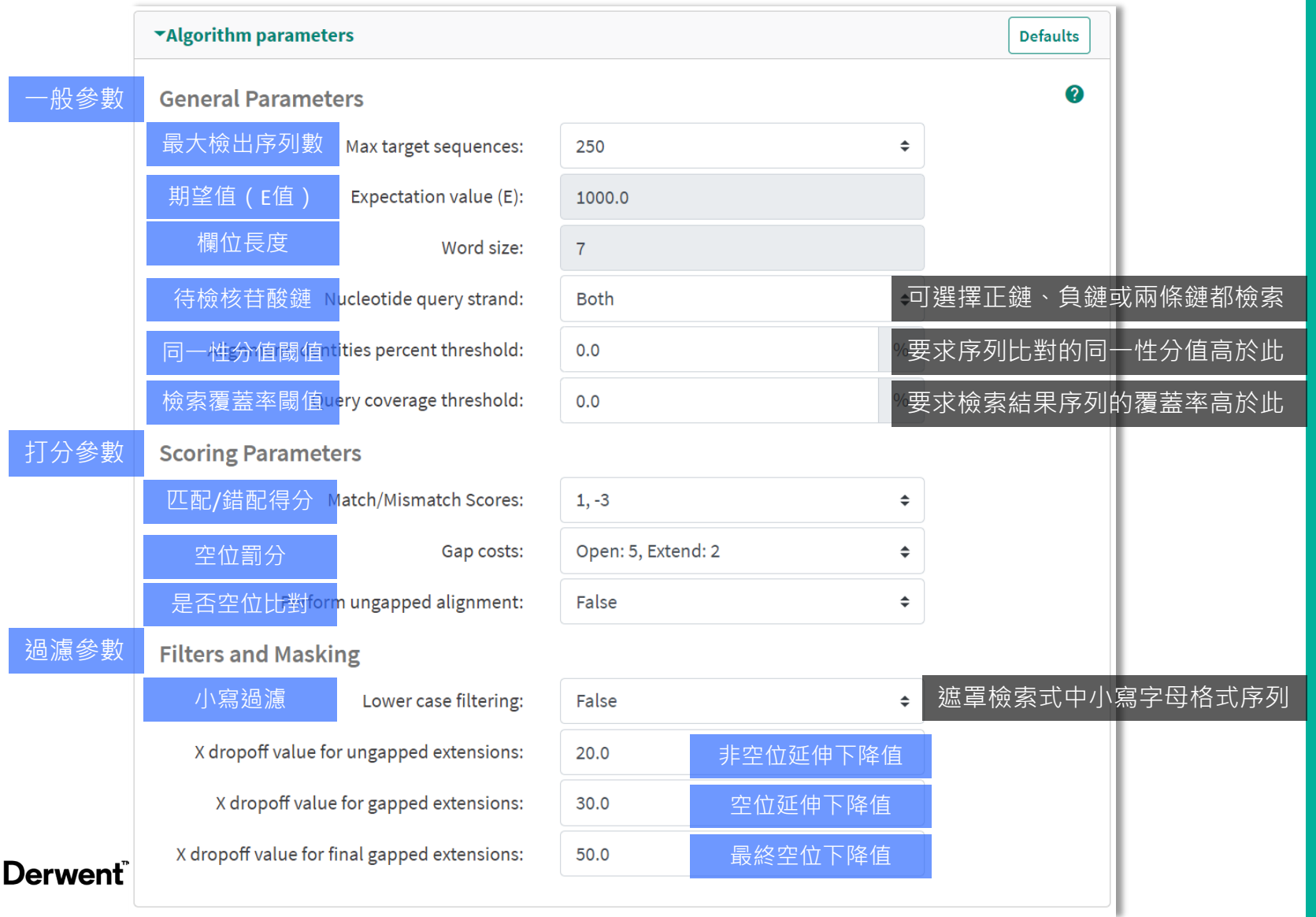
為了獲得理想的檢索結果並儘量減少避免不需要結果，可以在開始檢索之前指定一些參數和/或選擇所需的檢索選項。

通用參數設置在幾乎所有檢索方式中都會存在。包含：

- 目標序列的長度（上限為10w）
- 子序列長度

# 工作專案——檢索演算法參數

## BlastN檢索 演算法參數




### 附註說明

為了獲得理想的檢索結果並儘量減少避免不需要結果，可以在開始檢索之前指定一些參數和/或選擇所需的檢索選項。

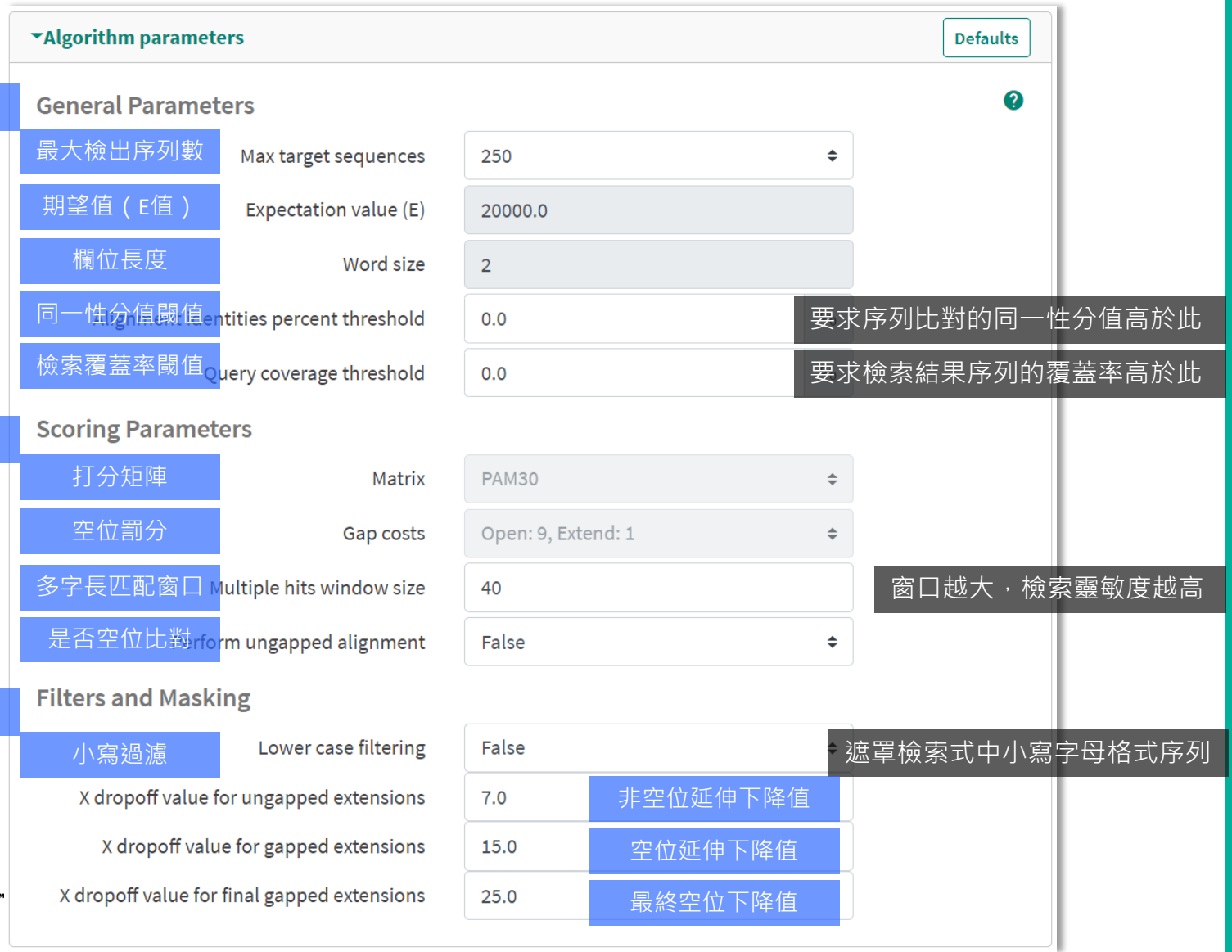
不同的檢索演算法對應的演算法參數也不盡相同。

左側為BlastN的演算法參數。

部分參數詳細解釋可參見附錄。

# 工作專案——檢索演算法參數

## BlastP/ TBlastN/ BlastX/ TBlastX/ MSS Protein 演算法參數




## 附註說明

為了獲得理想的檢索結果並儘量減少避免不需要結果，可以在開始檢索之前指定一些參數和/或選擇所需的檢索選項。

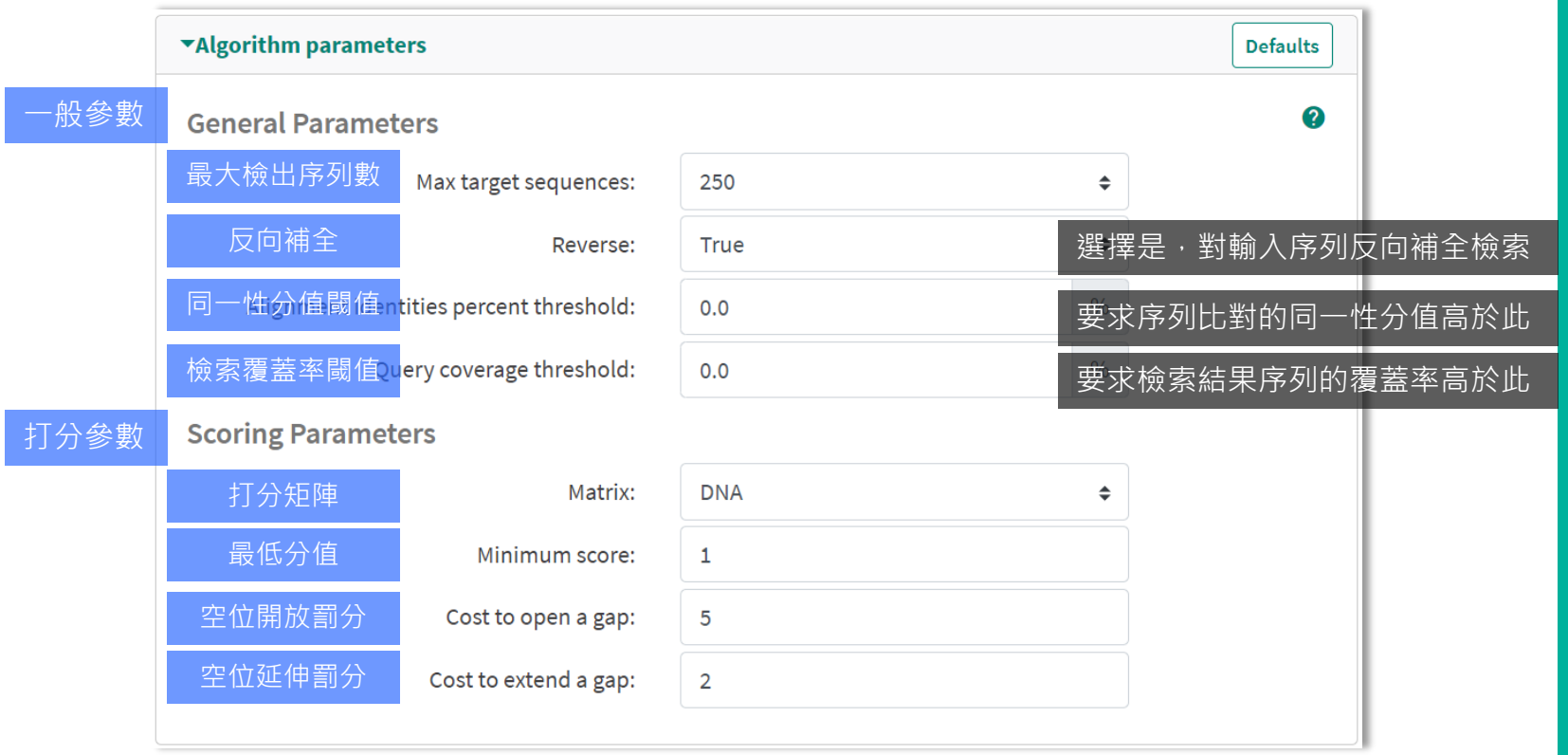
不同的檢索演算法對應的演算法參數也不盡相同。

左側為BlastP/ TBlastN/ BlastX/ TBlastX/ MSS Protein的演算法參數。

部分參數詳細解釋可參見附錄。

# 工作專案——檢索演算法參數

## Smith-Waterman N/ P、MSS Nucleotide演算法參數




## 附註說明

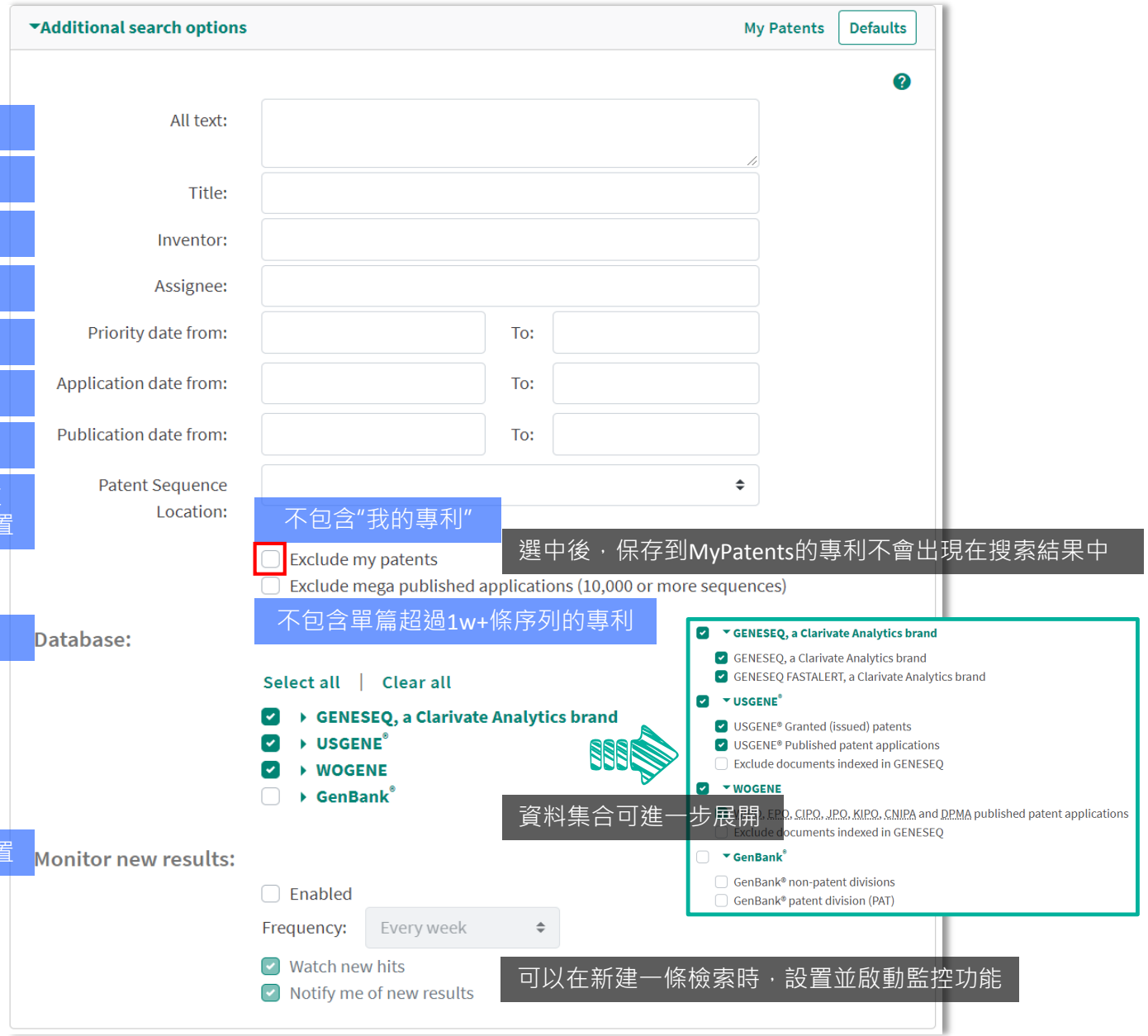
為了獲得理想的檢索結果並儘量減少避免不需要結果，可以在開始檢索之前指定一些參數和/或選擇所需的檢索選項。

不同的檢索演算法對應的演算法參數也不盡相同。

左側Smith-Waterman N/ P、MSS Nucleotide的演算法參數。

部分參數詳細解釋可參見附錄。

# 工作專案——其他檢索選項



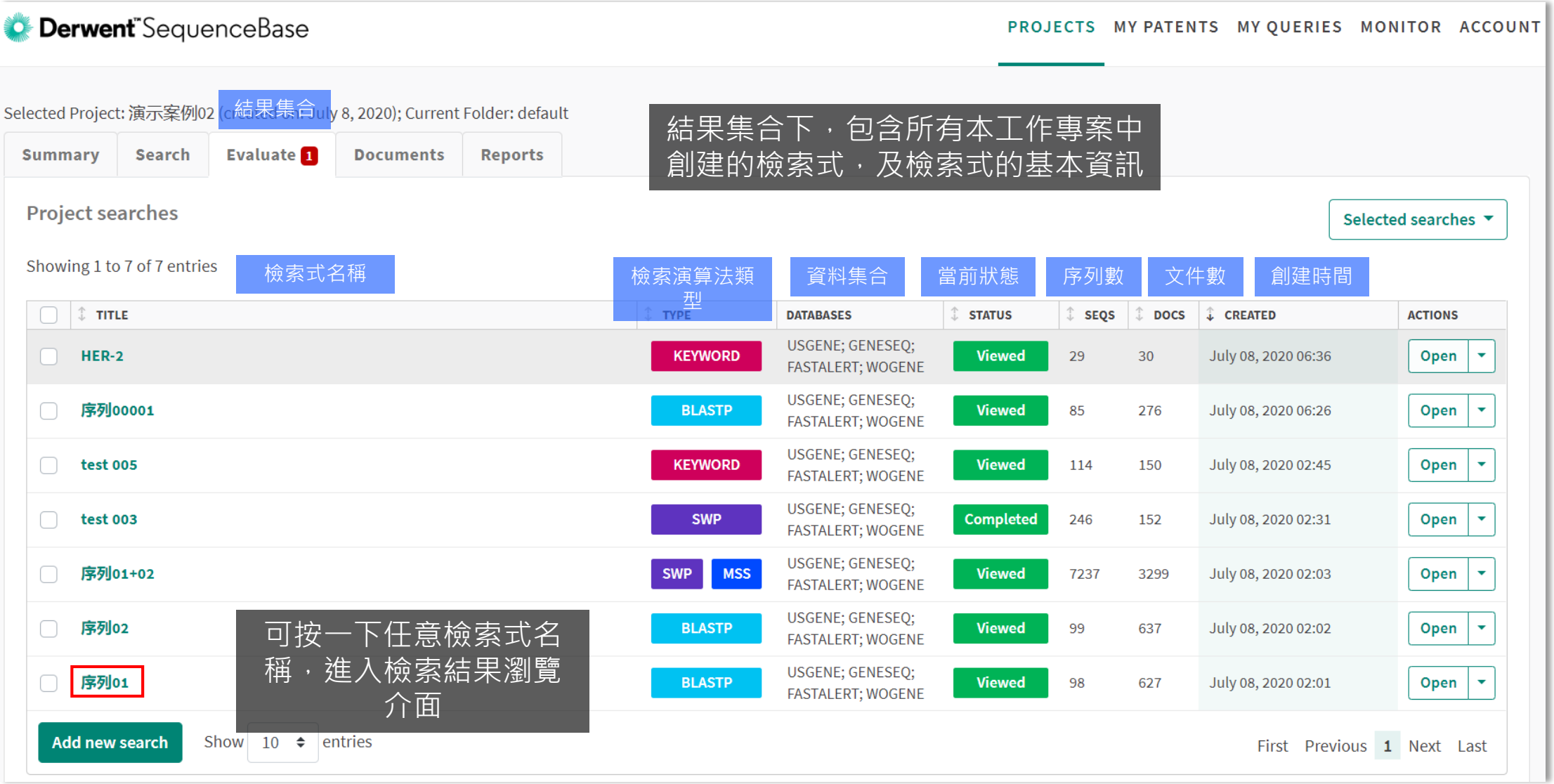

## 附註說明

其他檢索選項，可以說明我們在檢索序列的同時，增加限定條件，以使檢索結果更加精準。

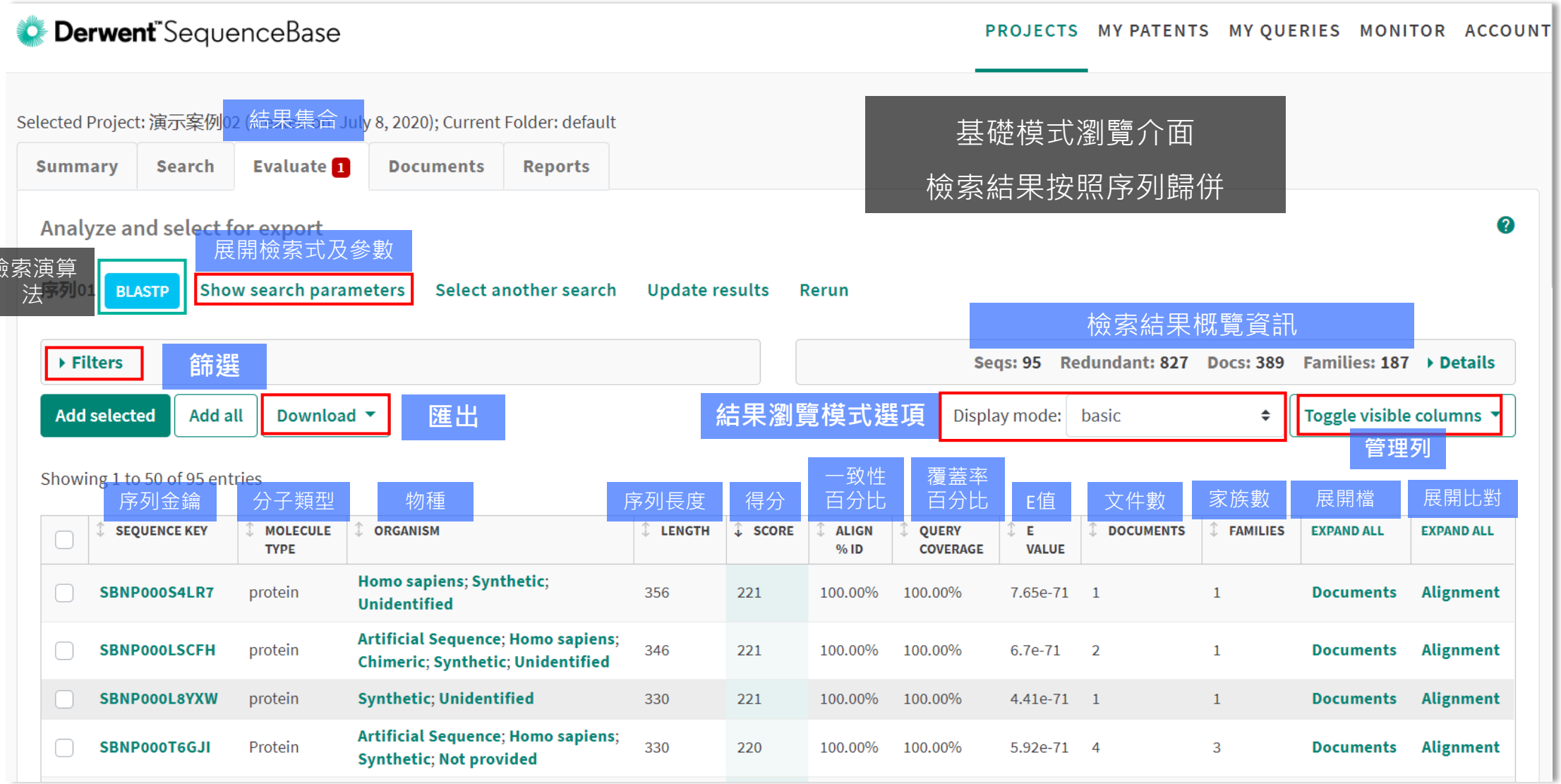
需要注意的是，在SequenceBase中，關鍵字均需輸入英文，不支援*等萬用字元。

此外，"全部文本"並非所有專利檔的全文，而是SequenceBase所有收錄文檔中的文本。

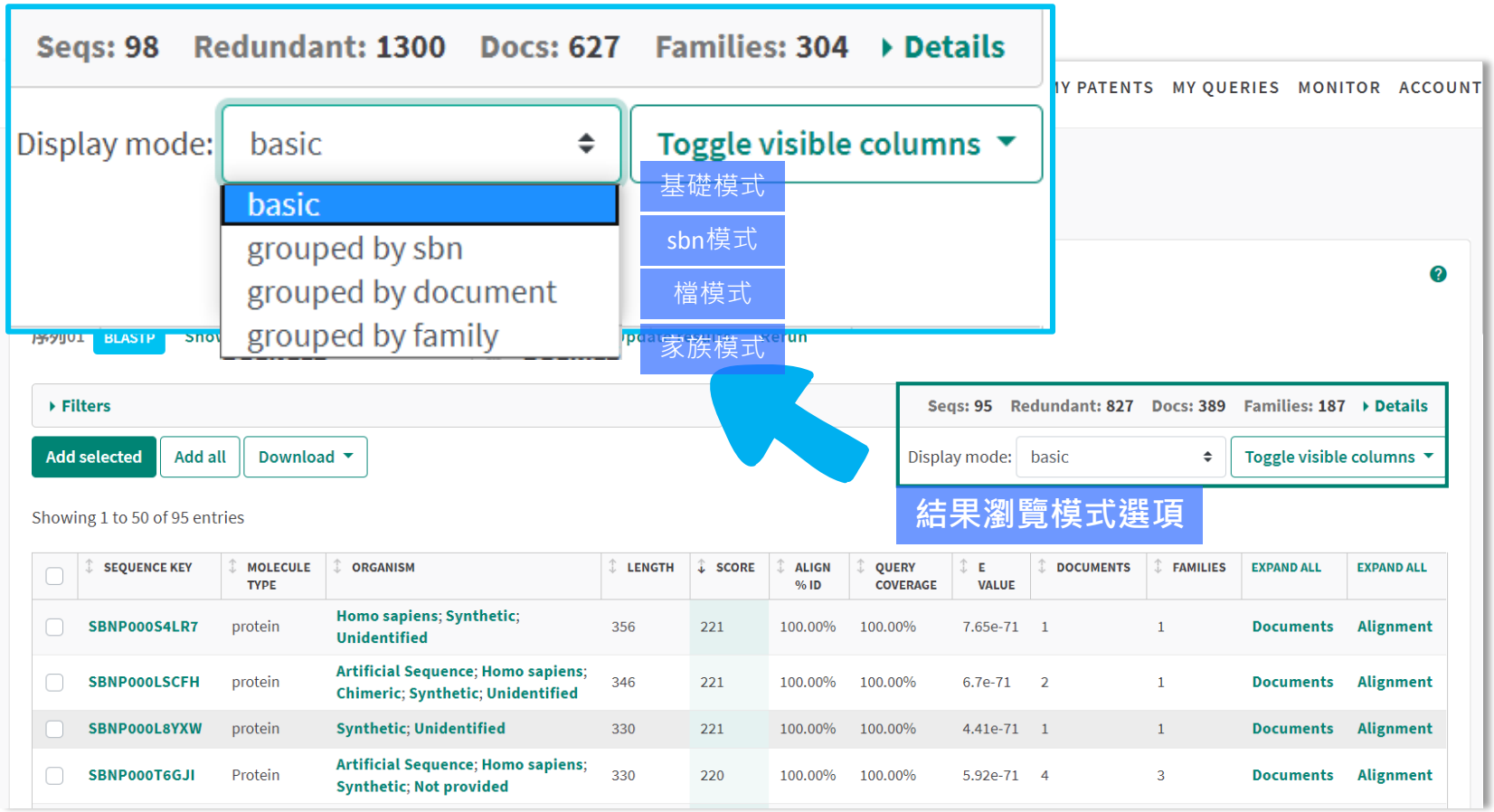
# 工作專案——檢索結果集合（Evaluate）



# 工作專案——檢索結果瀏覽（basic模式）



# 工作專案——檢索結果瀏覽——模式切換



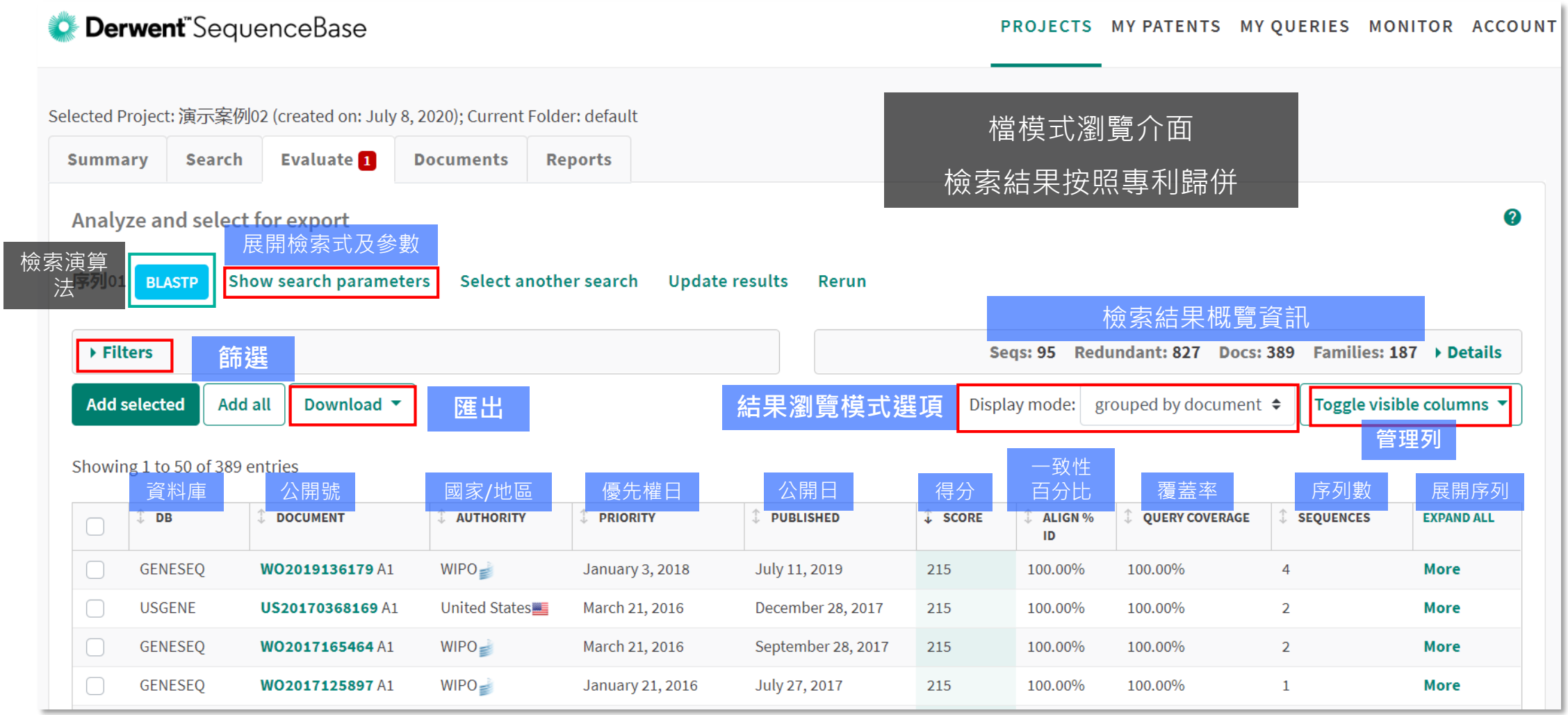

## 附註說明

檢索結果瀏覽包括四種模式：

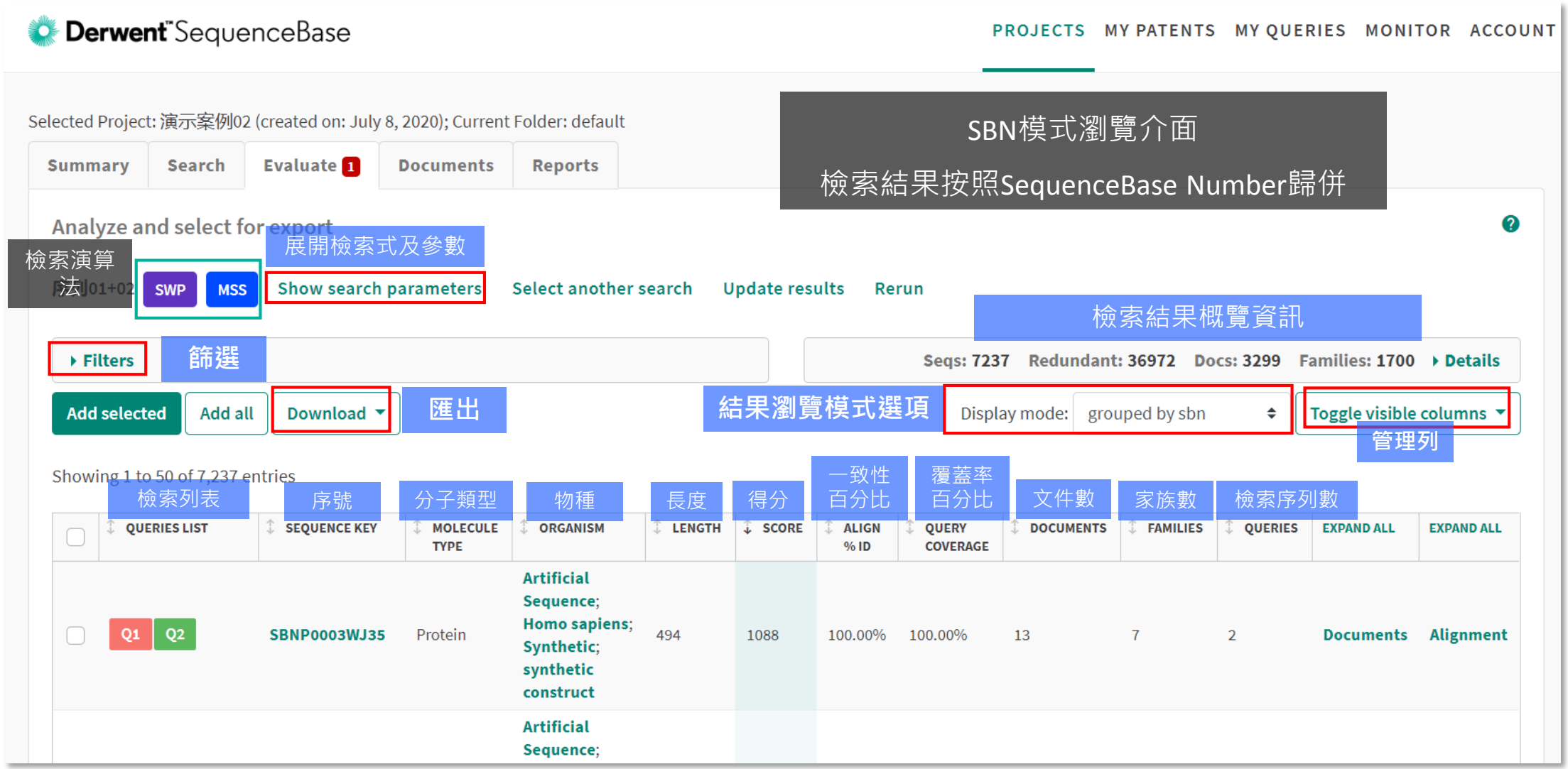
- Basic 基礎模式（檢索結果按照序列歸併）；
- Grouped by SBN SBN模式（檢索結果按照SequenceBase Number歸併）；
- Grouped by document 檔模式（檢索結果按照專利歸併）；
- Grouped by family 家族模式（專利按照家族歸併）''

**其中SBN模式僅在使用MSS檢索演算法的檢索結果中可以使用。**

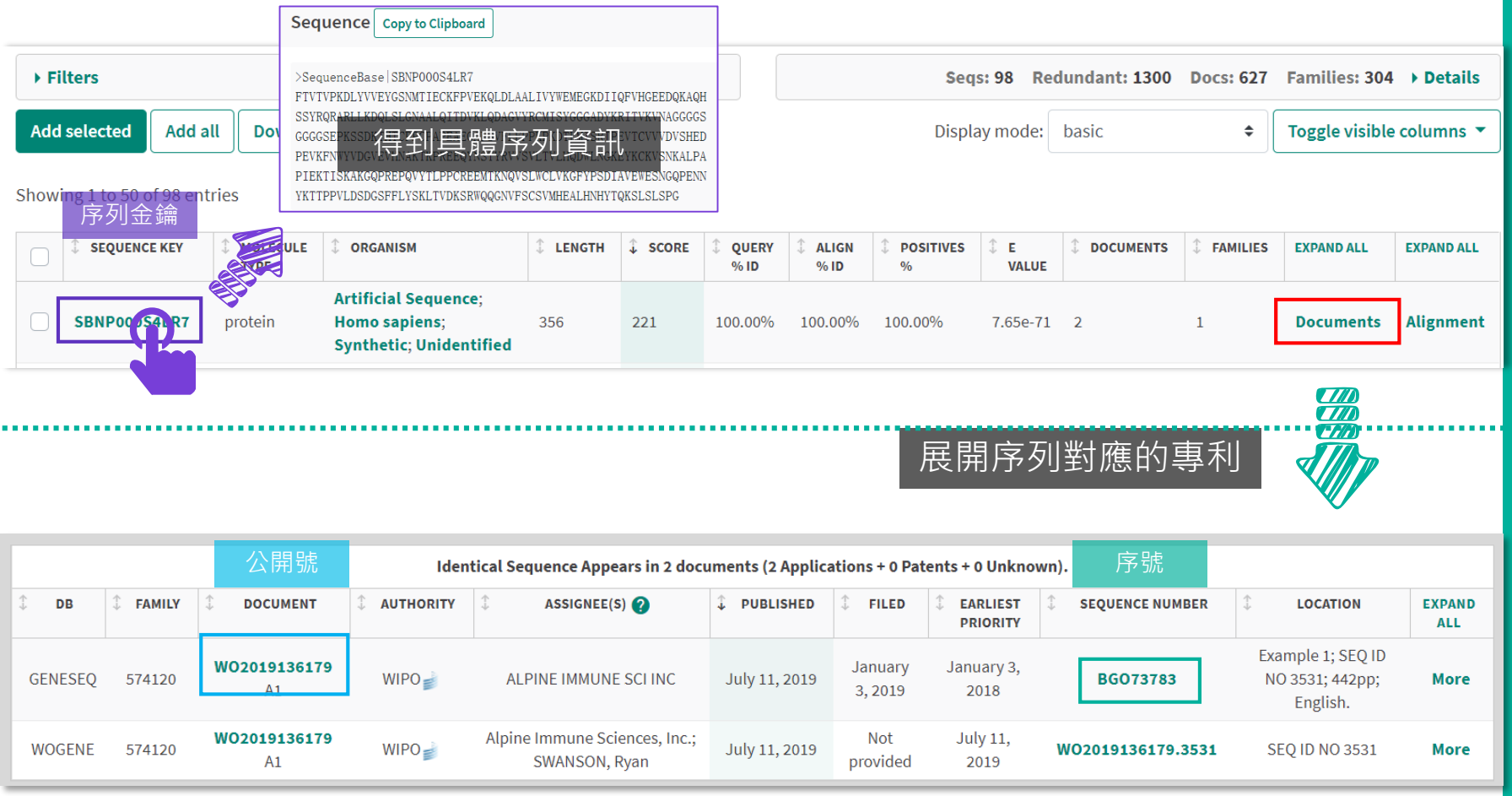
# 工作專案——檢索結果瀏覽（document模式）



# 工作專案——檢索結果瀏覽（SBN模式）



# 工作專案——檢索結果快速瀏覽




## 附註說明

結果瀏覽介面美觀直接

通過點擊Sequence KEY，可以得到具體序列資訊。

點擊物種，可以連結至Genbank的物種釋義。

點擊Document可以展開檢索結果中序列對應的專利。

# 工作專案——檢索結果快速瀏覽

展開序列對應的專利

公開號

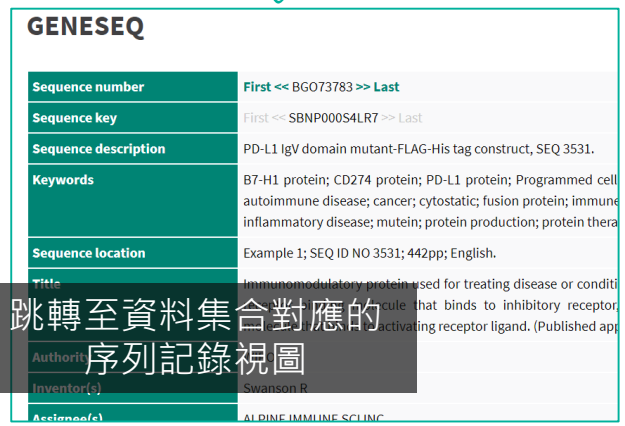
Identical Sequence Appears in 2 documents (2 Applications + 0 Patents + 0 Unknown).

序號

| DB | FAMILY | DOCUMENT | AUTHORITY | ASSIGNEE(S) | PUBLISHED | FILED | EARLIEST PRIORITY | SEQUENCE NUMBER | LOCATION | EXPAND ALL |
|---|---|---|---|---|---|---|---|---|---|---|
| GENESEQ | 574120 | WO2019136179 A1 | WIPO | ALPINE IMMUNE SCI INC | July 11, 2019 | January 3, 2019 | January 3, 2018 | BGO73783 | Example 1; SEQ ID NO 3531; 442pp; English. | More |
| WOGENE | 574120 | 2019136179 A1 | WIPO | Alpine Immune Sciences, Inc.; SWANSON, Ryan | July 11, 2019 | Not provided | July 11, 2019 | WO2019136179.3531 | SEQ ID NO 3531 | More |




跳轉至Derwent Innovation專利瀏覽介面




跳轉至資料集合對應的序列記錄視圖

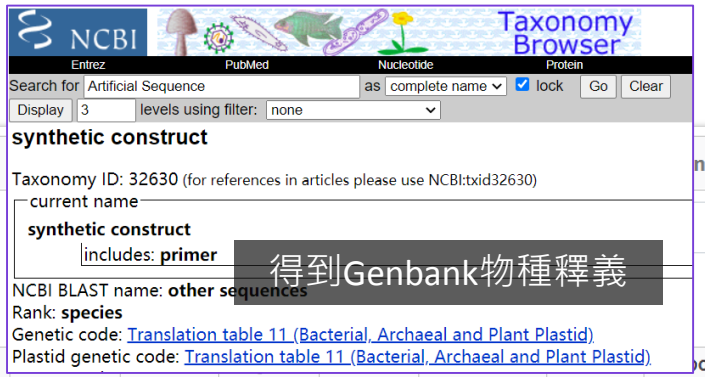
## 附註說明

展開檢索結果中序列對應的專利後

點擊專利的公開號，即可跳轉至Derwent Innovation的專利瀏覽介面，瀏覽專利的詳細資訊，並且可以下載專利pdf全文。

點擊序號，即可跳轉至SequenceBase的資料集合對應的序列記錄視圖中。

# 工作專案——檢索結果快速瀏覽




得到Genbank物種釋義

Filters

nt: 1300 Docs: 627 Families: 304 Details

Add selected | Add all | Download | Toggle visible columns

Showing 1 to 50 of 98 entries

物種

| SEQUENCE KEY | MOLECULE TYPE | ORGANISM | | | % ID | % ID | % | VALUE | DOCUMENTS | FAMILIES | EXPAND ALL | EXPAND ALL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SBNP000S4LR7 | protein | Artificial Sequence; Homo sapiens; Synthetic; Unidentified | 356 | 221 | 100.00% | 100.00% | 100.00% | 7.65e-71 | 2 | 1 | Documents | Alignment |

展開序列比對視圖

比對長度 | 得分 | E值 | 序列比對

Identical Sequence Appears in 2 documents (2 Applications + 0 Patents + 0 Unknown).

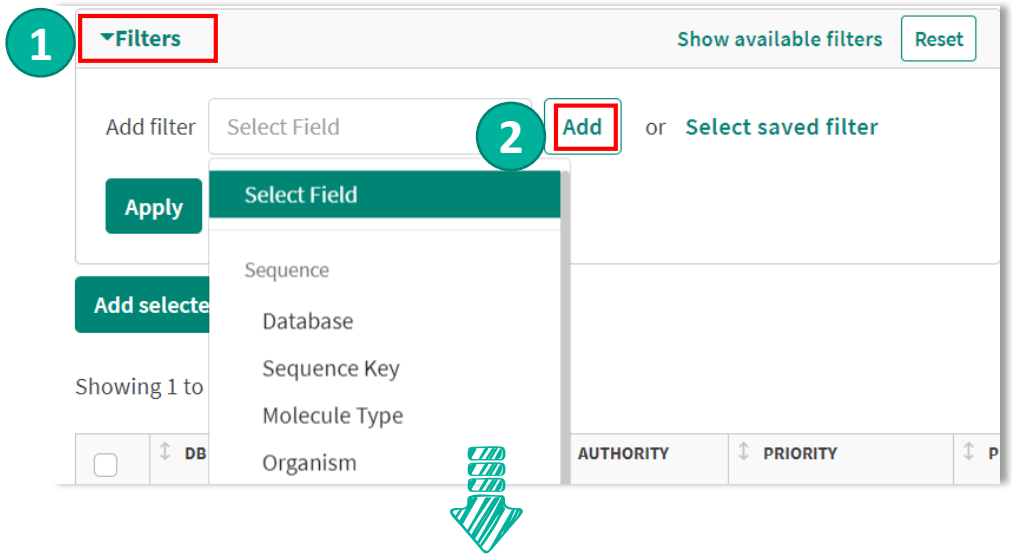
| ALIGNMENT LENGTH | SCORE | E VALUE | IDENTITIES | POSITIVES | GAPS | STRAND | ALIGNMENT |
|---|---|---|---|---|---|---|---|
| 103 | 221 bits (562) | 7.65e-71 | 103 | 103 | 0 | Plus/Minus | Query: 1 REPQVYTLPPCREEMTKNQVSLWCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGS 60<br>REPQVYTLPPCREEMTKNQVSLWCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGS<br>Sbjct: 254 REPQVYTLPPCREEMTKNQVSLWCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGS 313<br><br>Query: 61 FFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPG 103<br>FFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPG<br>Sbjct: 314 FFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPG 356 |

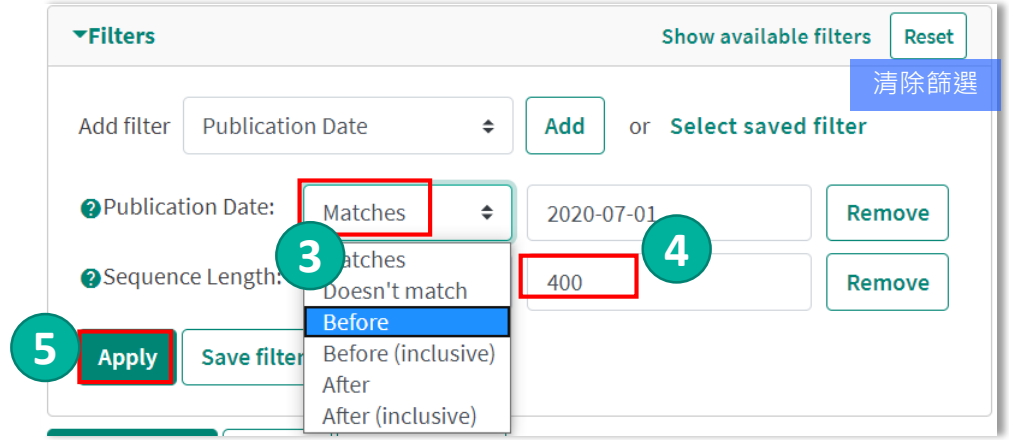
## 附註說明

結果瀏覽介面美觀直接

點擊Alignment可以展開序列比對詳情。

# 工作專案——檢索結果瀏覽——篩選




1 展開篩選器

2 選擇欄位並添加

3 選擇欄位匹配模式

4 增加欄位限定條件

5 篩選

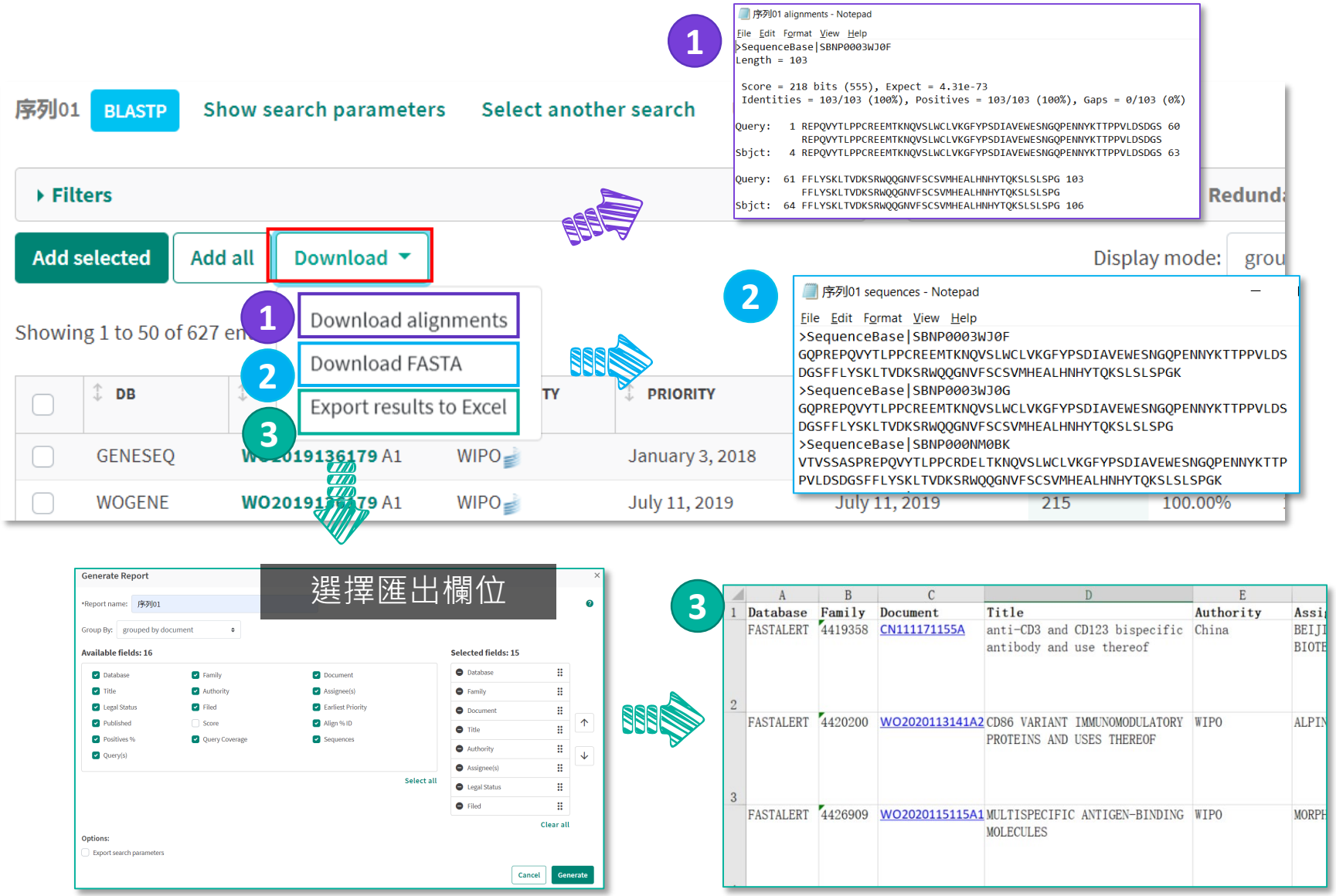
## 附註說明

篩選器可以幫助我們進一步限定檢索結果。

Derwent SequenceBase 的篩選器種類非常豐富，包括：

- 資料集
- 文本欄位/標題
- 專利權人/發明人
- 申請日/公開日/優先權日
- 序列在專利中的位置
- 序列長度/得分/E值/一致性百分比
- …… 

# 工作專案——檢索結果瀏覽——匯出




## 附註說明

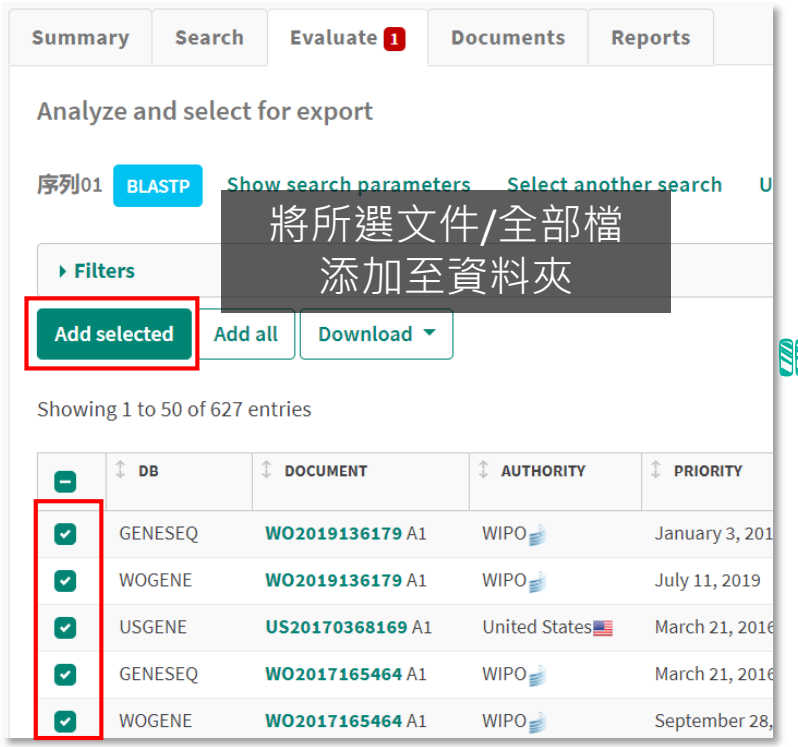
Derwent SequenceBase檢索結果支持三種形式的匯出：

1. 序列比對
2. FASTA格式序列
3. 檢索結果Excel列表

其中Excel格式中，可以通過點擊公開號連結，跳轉到Derwent Innovation中，查看專利詳情及下載專利pdf全文。

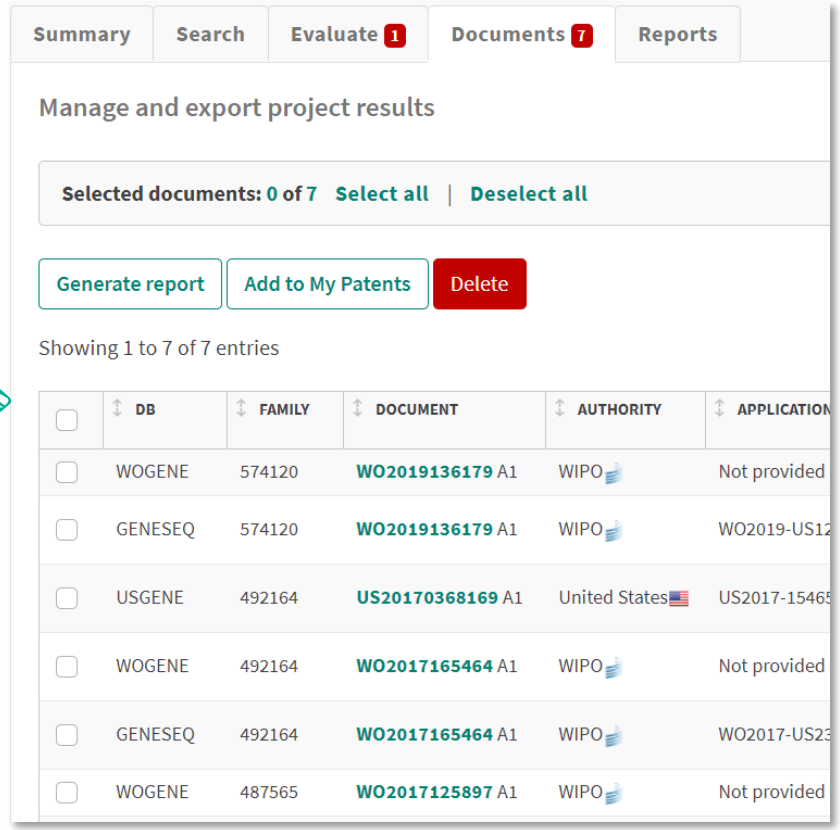
# 工作專案——文件

檢索結果瀏覽




將所選文件/全部檔添加至資料夾

文件

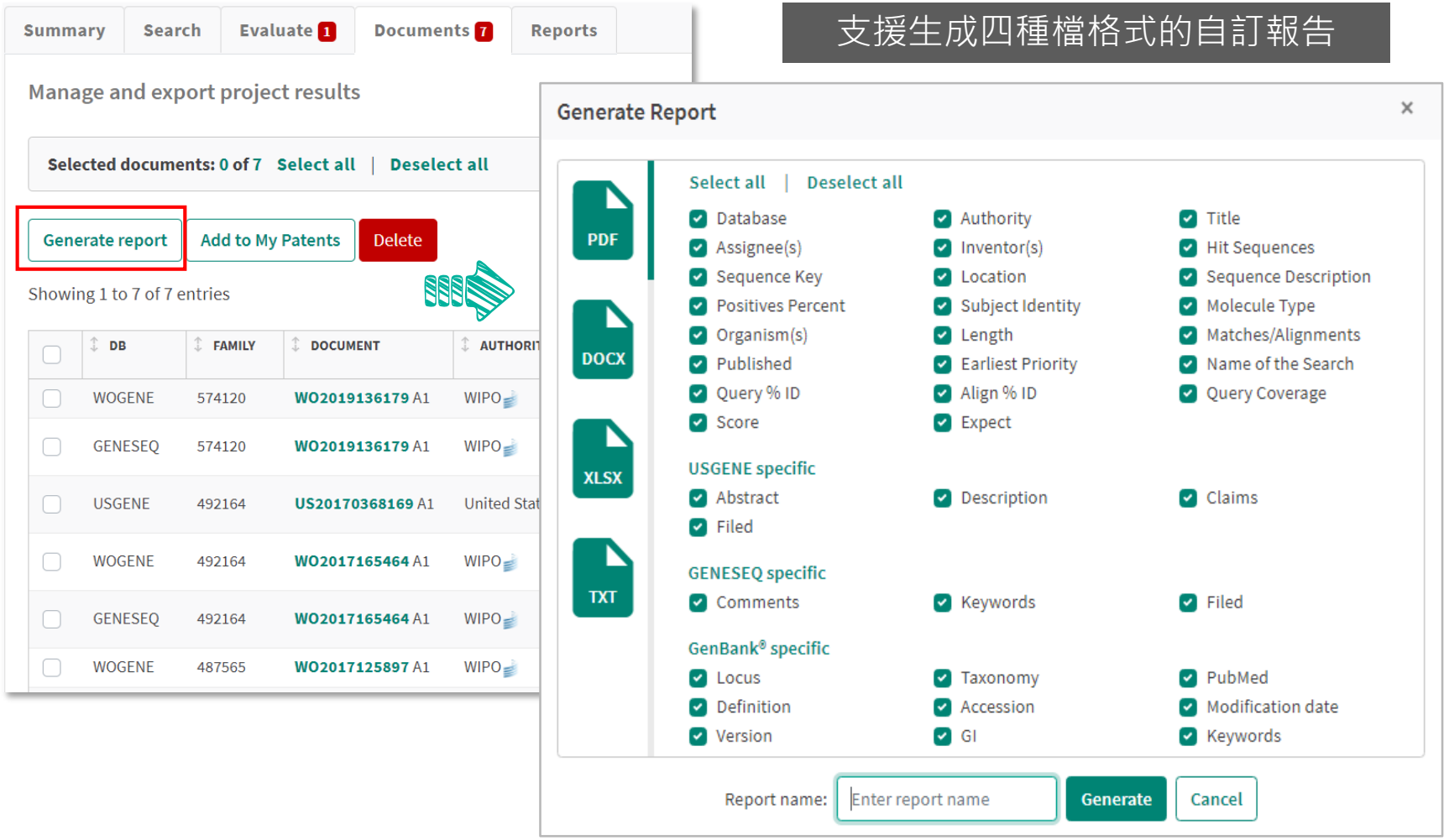



## 附註說明

Derwent SequenceBase支持將選中的檢索結果，或一個檢索式下的所有檢索結果添加至資料夾。

需要注意的是，當一個工作專案下的資料夾中的專利檔數量大於10000，該專案將會被關閉。

# 工作專案——文件




支援生成四種檔格式的自訂報告



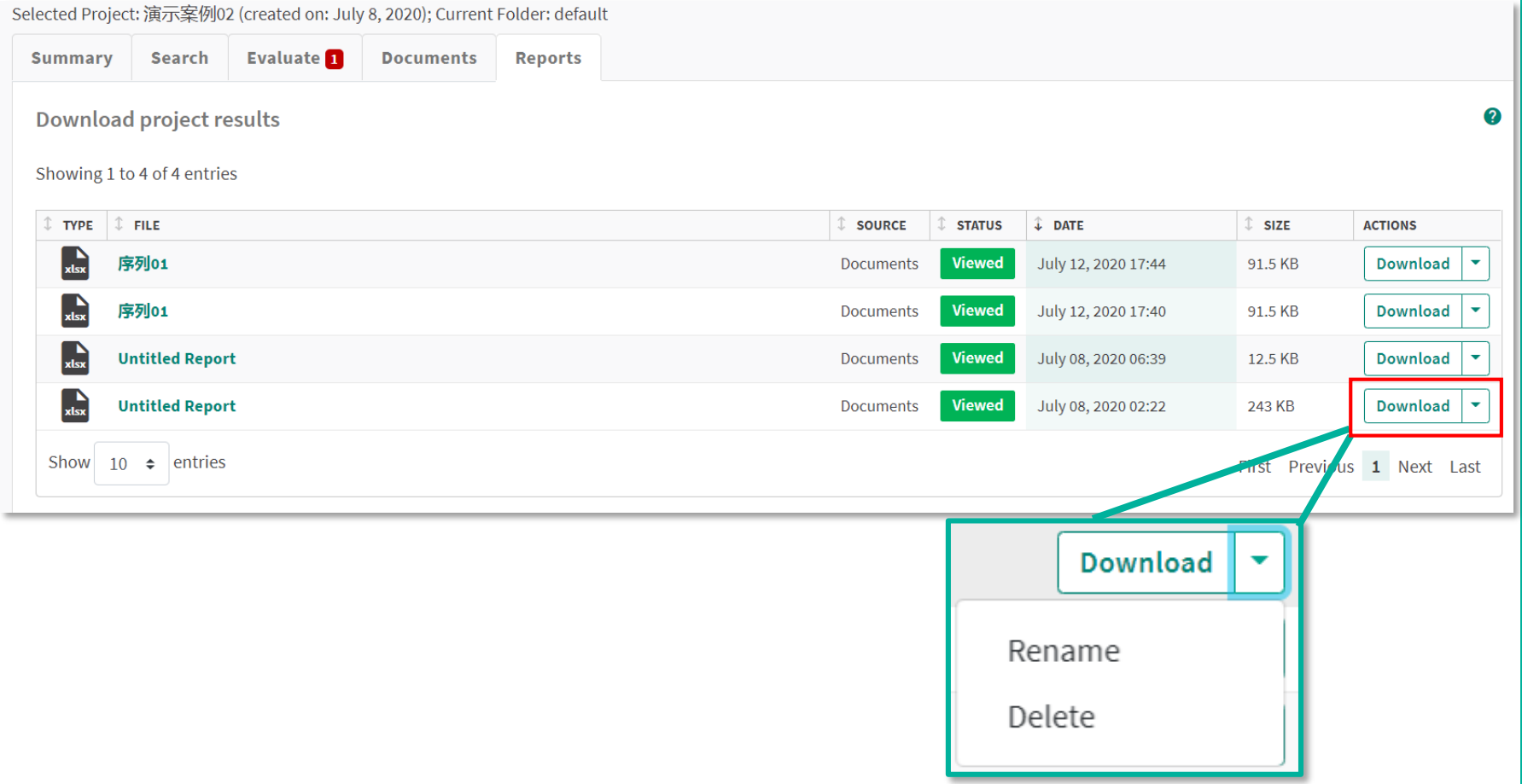

## 附註說明

資料夾的優勢在於：

可以將任意加入資料夾的檔，任選欄位，生成自訂報告。

報告支持PDF、DOCX、XLSX以及TXT四種格式。

# 工作專案——報告







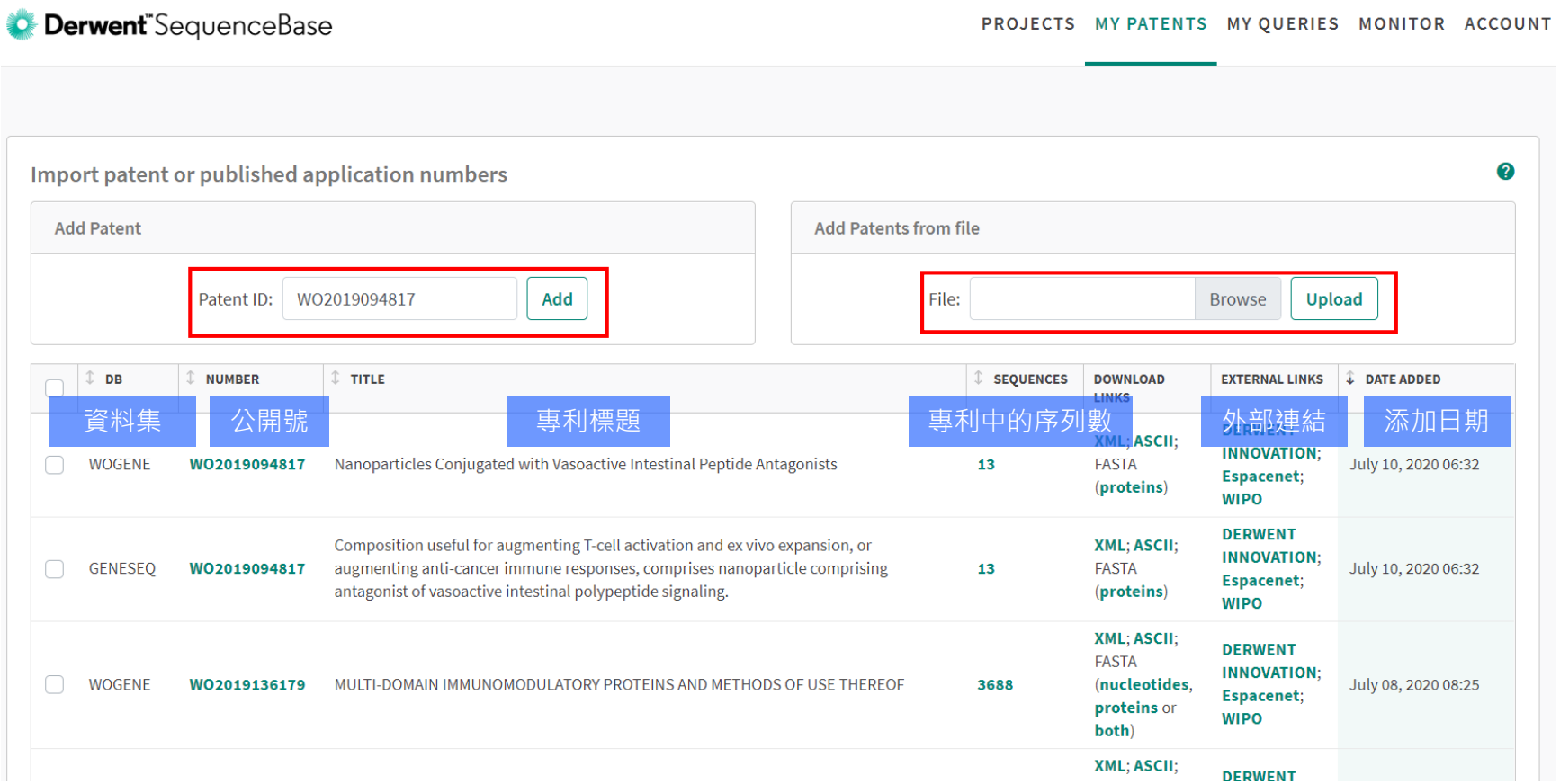
## 附註說明

報告介面

此頁面存儲您從檢索結果瀏覽介面的下載檔案，或通過文檔介面生成的所有報告。

在報告介面，您可以下載，重命名或刪除這些報告。

# My Patents 我的專利



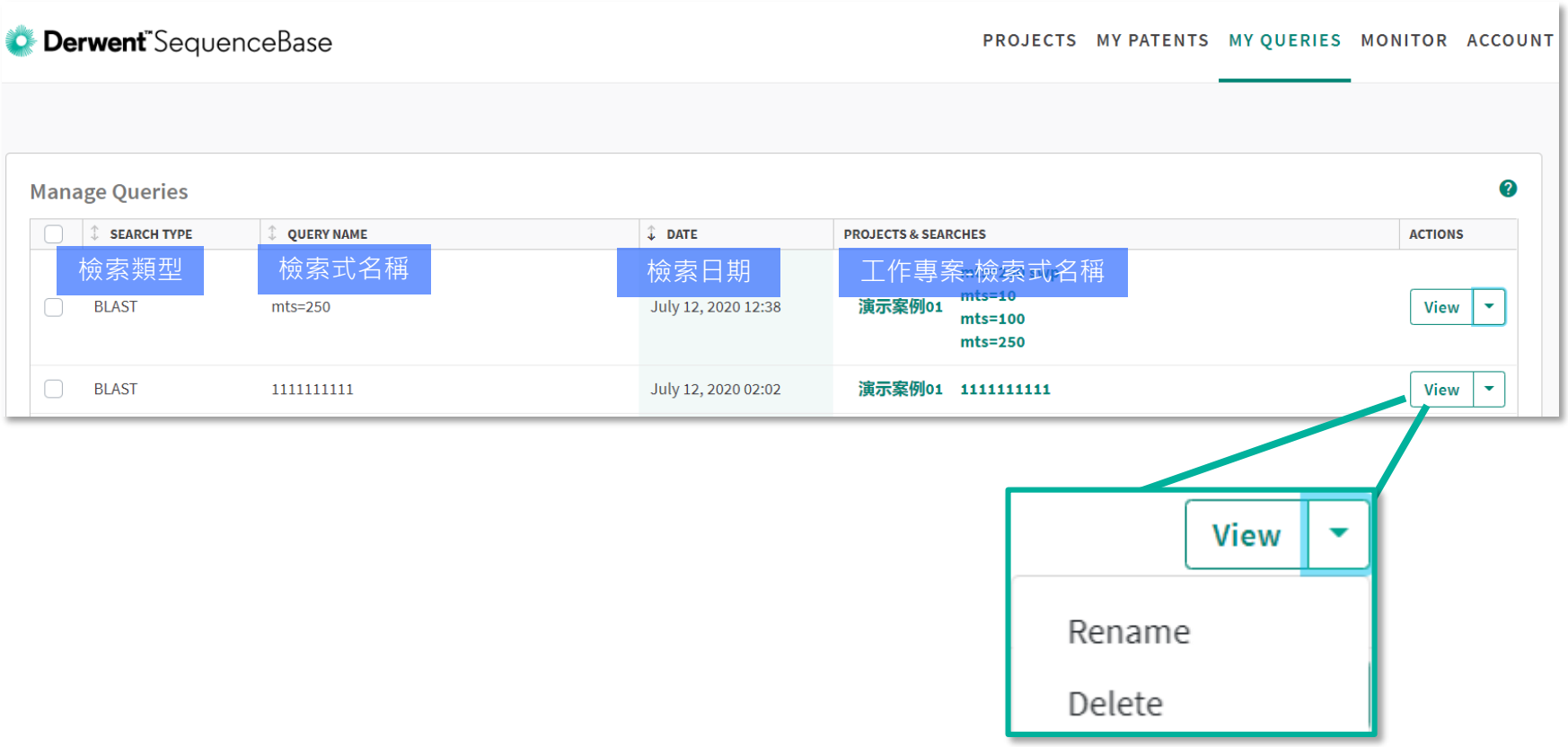

## 附註說明

My Patents

SequenceBase支持單件專利公開號輸入（捨棄公開號最後的專利文獻類型識別代碼），或檔批量上傳的形式添加MyPatents。

在這個介面下，我們可以非常方便的查看每件專利中所有序列的詳細資訊。

# My Queries 我的檢索







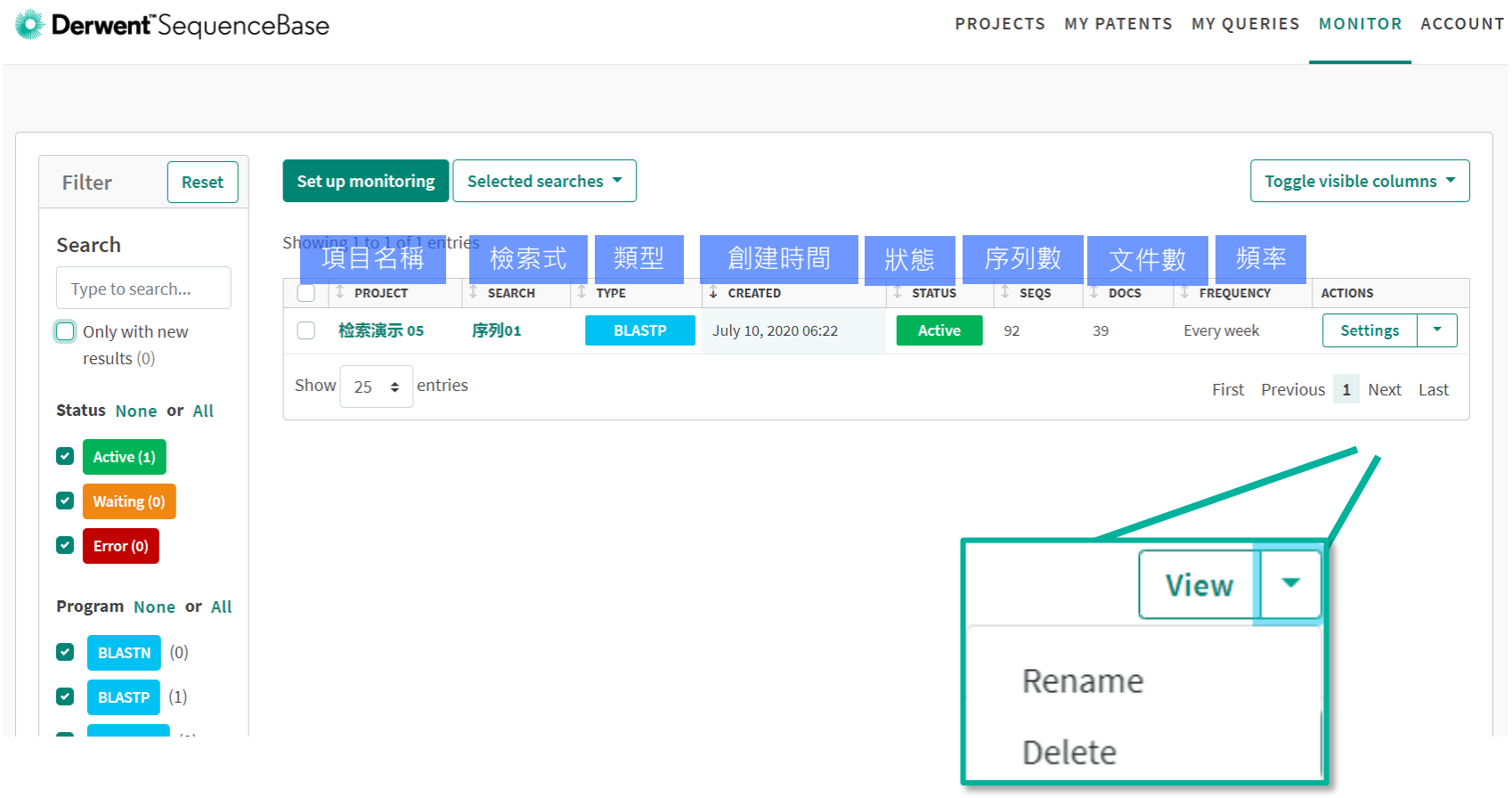
## 附註說明

My Queries

SequenceBase自動保存您的每一次檢索式。並可以將被檢索序列相同，但是工作專案/檢索式名稱不同的檢索式自動歸併。

在這個介面下，我們可以非常方便的瀏覽所有的檢索記錄，更改名稱或刪除。

# Monitor 監控







## 附註說明

Monitor 監控
SequenceBase支援檢索式自動監控功能。

資料庫將不斷使用新文檔和新序列進行更新。 監視功能使使用者可以重新開機精確搜索，以查找新序列和與提交序列匹配的文檔。
監視將發現新序列，冗余序列，文檔和族。
它允許您：
-觀看新專輯
-跟蹤您的更新歷史記錄–在更新之前查看搜索結果，即可以訪問特定日期的搜索結果資訊（評估→篩檢程式→更新歷史記錄）
-僅查看新結果–選擇僅查看最新更新後出現的新結果
-設置自動更新（監控→設置）
-在出現新結果時設置電子郵件通知（監控→設置）

# 附錄 I

## 序列檢索演算法參數解讀

# 演算法參數解讀

## Expectation value (E)

## 期望值 (E值)

- ***E* 值**：是BLAST的一個非常重要的**統計學**指標。
- **理解*E* 值的前提：**

**1**、BLAST比對結果中，得分高、非常相似的兩條序列 有可能是隨機造成的，無生物學意義（不同源），但被BLAST報告出來。 反之，得分低、不相似的兩條序列也可能不是隨機的，有生物學意義（同源），但被BLAST漏報。

**2**、因此，BLAST輸出的比對結果就會出現兩種疏漏，**假陽性**（上述前一種情況）和**假陰性**（上述後一種情況）。且兩種情況**此消彼漲**。即，當我們增加BLAST輸出的搜索結果時，**假陰性**會減少（但**假陽性**增加）；反之亦然。

3、顯然，BLAST在輸出比對的結果時，需要在減少假陽性（盡可能減少隨機比對的輸出)與減少假陰性（盡可能不要漏報）之間尋找一個平衡。

4、E值正是用來解決這一平衡的統計學指標。

# 演算法參數解讀

## Expectation value (E)

## 期望值 (E值)

***E* 值的定義**：一次搜索中（特定的資料庫、打分矩陣及相關參數），期望發生的比對得分等於或大於某一分數的隨機比對數目。

- E = 10，我們期望在BLAST搜索結果中，比對得分等於或大於某一分數，屬於隨機匹配的比對數目為10個。
- E = 1 ……
- E = 0.01 ……

從BLAST實踐來講，如果我們將E值調低，則BLAST返回的搜索結果就會變少（假陽性減少，但假陰性增加）；反之，將E值調高，則返回的搜索結果就會增多（假陰性減少，但假陽性增加）。

# 演算法參數解讀

**Word size**

**欄位長度**

**Matrix**

**打分矩陣**

**Word size**是BLAST演算法中另一個參數，對於短序列的檢索很重要。

BLAST演算法使用欄位來實現相似性區域的匹配。我們待查詢序列的長度必須至少是欄位長度的兩倍（例如，如果您查詢的是4個殘基的蛋白質序列，則字長應最多為2）。

請注意，欄位長度大小越小，獲得檢索結果越多，檢索所需的時間就越長。

**Matrix**

序列比對的打分矩陣，BlastP 提供5種打分矩陣：

PAM30、 PAM70、 BLOSUM45、 BLOSUM62、BLOSUM80。

BlastN無此選項

# 演算法參數解讀

## Match/Mismatch Scores

## 匹配/錯配得分

### Match Score（ BlastN only ）

在用BlastN演算法檢索序列時，比對上的每個核酸位點的得分，預設值是1。

### Mismatch Score（ BlastN only ）

在用blastn程式搜索序列時，設定核酸錯配的罰分，預設值是-3。

## Gap Costs

## 空位罰分

### Gap Costs

OPEN 空位開放　　Extend 空位延伸

基因在進化過程中往往會產生鹼基的插入或缺失。這樣，在進行序列比對時，為了更好地反映序列的相似性，也就必須考慮引入空位並進行罰分以控制空位插入的合理性。

# 演算法參數解讀

**Lower case filtering**

小寫過濾

**Multiple hits window size**

多字長匹配窗口

## Lower case filtering（小寫過濾）

將檢索式中小寫字母格式的序列遮罩，該功能可以濾除FASTA格式序列中的小寫字母。

在蛋白質序列中，一些不準確區域通常用小寫字母表示。

## Multiple hits window size

指定多字長匹配的視窗大小，視窗越大，BLAST檢索的靈敏度越高。

# 附錄 II

## SequenceBase資料庫資料來源

# 資料來源

## SequenceBase資料庫四大資料來源

- GENESEQ
- WOGENE
- USGENE
- GenBank

## GENESEQ

GENESEQ™ 是由Clarivate創建的生物序列專利資料庫，包含WIPO，USPTO，EPO，JPO，KIPO，IP Inda和CNIPA等56 個國家/地區專利頒發機構的生物序列（DNA、RNA 及蛋白質序列）專利。該資料庫的記錄可追溯到1981年，每個記錄都包含人工加工的標準化標題、書目專案資訊，英文摘要以及人工捕獲的序列資訊。GENESEQ每兩周更新一次。**截至2020年7月**，GENESEQ及其伴隨資料庫GENESEQ FASTAlert中有150,815,798個序列記錄，摘自561,305 個德溫特世界專利索引（DWPI）基本專利。

GENESEQ™ 的人工加工欄位重點標識了每件專利中特定序列記載的位置以及其生物學意義，可有效的減少智慧財產權及生物技術專家檢索和解讀專利文本的時間，更專注於判斷技術的可專利性、識別潛在侵權風險以及追蹤競爭對手動態。

# 資料來源

## SequenceBase資料庫四大資料來源

- GENESEQ
- WOGENE
- USGENE
- GenBank

## USGENE

USGENE®是美國專利商標局（USPTO）的序列專利資料庫，包含USPTO已公開專利申請和已公告專利的所有可用DNA，RNA和蛋白質序列的帶注釋集合。該資料庫的記錄可追溯到1981年，每週二和週四更新。**截至2020年7月**，USGENE中有256,327,097個序列記錄，這些記錄來自207,614個已公開的專利申請和134,807個已公告的專利。

## WOGENE

WOGENE是全球序列專利資料庫，包含WIPO，EPO，CIPO，JPO，KIPO，CNIPA和DPMA已公開專利申請中所有可用DNA，RNA和蛋白質序列的帶注釋集合 。該資料庫的記錄可追溯到1979年， WIPO的資料在公開當天即可查詢。**截至2020年7月**，WOGENE中有111,646,214個序列記錄，摘自352,287個已公開的專利申請。

# 資料來源

## SequenceBase資料庫四大資料來源

- GENESEQ
- WOGENE
- USGENE
- GenBank

## GenBank

GenBank®是美國國立衛生研究院（NIH）的基因序列資料庫，收錄所有公開的 DNA 序列，並且附帶注釋。GenBank®屬於國際核苷酸序列資料庫合作聯盟 (INSDC)，該組織成員包括日本 DNA 資料庫 (DDBJ)、歐洲核苷酸檔案館 (ENA) 和 NCBI 的 GenBank。

GenBank擁有超過303,339,826百萬個序列記錄，是現有技術中非專利文獻的寶貴來源，可提供最新和最廣泛的DNA序列資訊。

GenBank資料可通過Derwent SequenceBase獲得並檢索。

# 科睿唯安智慧財產解決方案

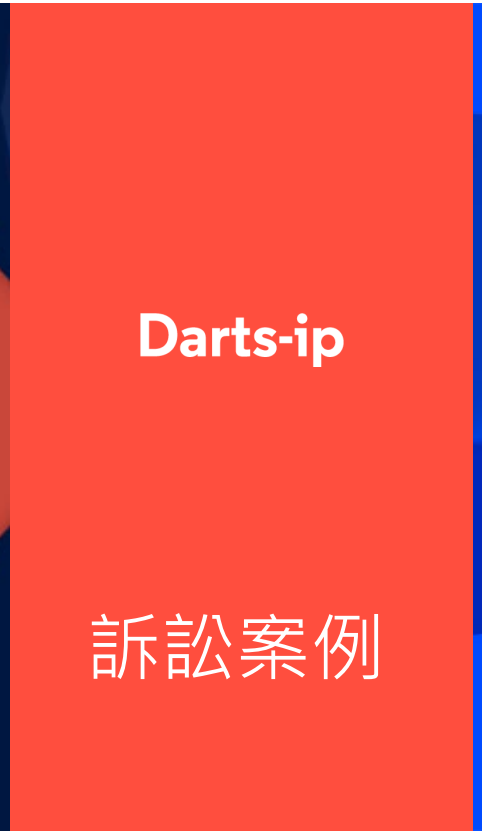
專利、商標、訴訟、網域名稱及工業標準，科睿唯安的智慧財產解決方案都是各領域的領導者。

我們有為全球的創新公司提供服務的豐富經驗，是助力創新機構發展的最佳夥伴。















# 科睿唯安智慧財產解決方案的價值

我們幫助智慧財產專業人員、研發人員、市場人員，運用資料的力量大幅提高生產力，讓他們將時間及資源集中為客戶和公司創造價值。



## 數據

- 9,000萬條高品質商標資料
- 1億多件專利資料及4,300萬件深加工改寫
- 超過300萬件商標訴訟和150萬件專利訴訟案例
- 55萬個標準資料
- 管理全球130萬個網域名稱



## 專家

- 300多位元具有智慧財產背景的資料團隊，平均擁有18年以上行業經驗
- 熟悉專利、商標、標準和案例資料
- 自1999年以來ICANN認可的網域名稱註冊商，管理的網域名稱每月有120億網站訪客



## 客戶體驗

- AI驅動的獨家商標圖像檢索技術
- DWPI德溫特世界專利索引能為使用者帶來極大的資訊價值
- 為基因序列獨特需求提供最頂尖的基因序列檢索平臺與資料
- 提供高度安全的網域名稱獲取管道

# 全球客戶



全球40多個專利局採用Derwent進行專利審查



全球有超過40萬個創新者或法律專業人士，在使用Derwent Innovation尋找技術資訊或支援法律訴訟



全球10大最有價值的品牌*有9個採用CompuMark來保護他們的商標

*Interbrand 2019 Best Global Brands



過去十年，我們為3,500多個商標客戶做過客製化服務



全球十大流量最高的網站*，信任MarkMonitor為其註冊及管理網域名稱組合

*Alexa



全球有13萬個對工業標準有需求的用戶選擇使用Techstreet

在科睿唯安，

我們深信人類的創造力

可以改變世界與未來。

我們的願景是改善全球

促進創新和保護創新的

方式



Clarivate™ 科睿唯安™

# 認識科睿唯安



在整個創新生命週期中，我們向各地的
創新人才提供重要關鍵的數據、資訊、
工作流解決方案和深厚專業知識。

我們提供可靠的見解和分析，
是客戶加快創新步伐過程中
不可或缺的合作夥伴。



# 全球42個國家和地區

# 超過4,300名員工



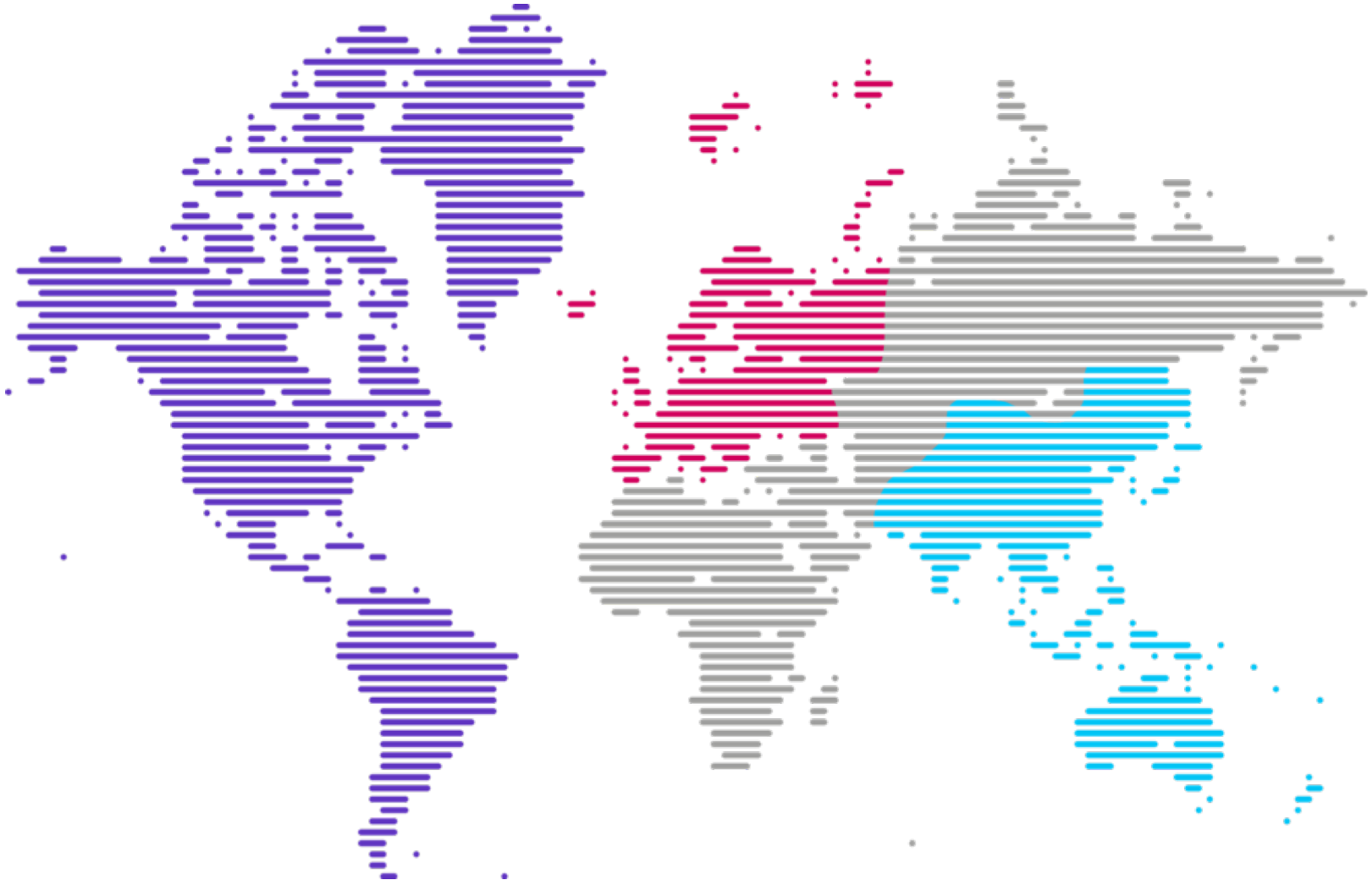
**37%**
美洲



**27%**
歐洲



**36%**
亞太



# 由950多名專家組成的內容編輯團隊

一半以上擁有碩士或博士學位，內容編輯團隊負責對非結構化資料進行清理、分析和分類等深加工作業，以確保我們提供的資料品質，優化用戶的體驗。



# 超過150年的專業歷史傳承

**1864**
Zoological Society of London 及 British Museum 建立 Zoological Records動物生物學資料庫

**1926**
Society of Bacteriologists 及 Botanical Society 建立BIOSIS生物學資料庫

**1955**
Eugene Garfield 博士發明引文索引



**1957**
成立 Institute for Scientific Information (ISI)公司



**1963**
專利家族之父 Monty Hyams開始製作發行 Derwent專利摘要



**1986**
湯森公司收購商標檢索公司 Thomson & Thomson 及 CompuMark

**1997**
推出 Web of Science

**2003**
湯森公司收購 Techstreet

**2011**
推出 Cortellis

**2012**
湯森公司收購 MarkMonitor

**2016**
智慧財產與科學事業部從湯森路透獨立，成立科睿唯安 (Clarivate Analytics)

Clarivate Analytics

**2017**
收購 Publons

publons

**2018**
收購 Kopernio

Kopernio

**2018**
收購 TrademarkVision

TrademarkVision

**2019**
與Churchill 公司合併，科睿唯安 (Clarivate) 正式在紐交所上市



**2019**
收購 SequenceBase

SequenceBase

# 科睿唯安深獲世界頂尖機構的信賴

**18,000** 家客戶

**遍佈全球180多個國家和地區**

**包含大學、非營利組織、科研機構、學術期刊出版商、民營企業、政府機構、法律事務所**



**50**

全球頂級生物製藥公司



**40+**

專利局



**50**

全球最大的律師事務所



**7,000+**

頂尖的政府/學術科研機構及重視創新研究的企業



**10**

全球十大流量最高的網域名

# 全球領導團隊



**Jerre Stead**
Executive Chairman
and CEO



**Mukhtar Ahmed**
President,
Science Group



**Andrea Degutis**
Chief Human
Resources Officer



**Richard Hanks**
Chief Financial Officer



**Steve Hartman**
General Counsel and
Global Head of Corporate
Development



**Randy Harvey**
Chief Technology Officer



**David Liu**
Managing Director,
Asia Pacific



**Julia Mair**
Chief Marketing Officer



**Mike Morhardt**
Executive Vice President,
Commercial Performance
and Operations



**Armughan Rafat**
Chief Data Officer



**Jeff Roy**
President, IP Group



**Kathy Sullivan**
Senior Vice President,
Revenue Operations

Clarivate
Derwent™

# 聯繫我們

臺北市信義區松智路1號11樓

0080 149 1138

https://clarivate.com.tw

https://clarivate.com/derwent/

歡迎關注我們的粉絲團

© 2020 Clarivate. All rights reserved. Republication or redistribution of Clarivate content, including by framing or similar means, is prohibited without the prior written consent of Clarivate. Clarivate and its logo, as well as all other trademarks used herein are trademarks of their respective owners and used under license.