**The Village Database**

*A Resource for Chinese American Genealogy Research*

---

**RESEARCH FILES**

*by* Henry Tom

**UNTIL 1965,** over 90 percent of all Chinese in North America originated from Guangdong Province (廣東 Kwangtung), which includes Zhongshan County (中山 Chungshan), Sanyi District (三邑 Sam Yup), a three-county region composed of Nanhai County (南海 Nanhoi), Panyu County (蒲崗 Panyu), Shunde County (順德 Shuntak), and Siyi District (四邑 Say Yap or Sze Yup), a four-county region that includes Xinhu County (新會 Sunwui or Sunwei), Taishan County (台山 Toisan or Toishan), Kaiping County (開平 Hoiping), and Enping (恩平 Onping or Yanping). Within Sanyi, Taishan County accounts for over half of the 90 percent of this migration. As a retired U.S. federal geographer and cartographer, I have spent the past sixteen years helping Chinese Americans research their ancestors from these regions using the Village Database directory. I have used this invaluable resource to conduct over 3,000 queries, of which over 85 percent have found successful matches. Currently, no comparable database of surnames and village name exists for Chinese originating outside the Siyi region. There are approximately 50 million people of Chinese ancestry across the world, many originating from Guangdong Province, so the Village Database is useful globally as well as by those living in the United States. The following narrative provides an explanation of how to use the Village Database to find the ancestral villages of Chinese people originating from Guangdong province without having to be Chinese-literate or a geographer. My hope is that this article will help future researchers use this resource to learn more about their own family histories.

---

**THE VILLAGE DATABASE AND WHY IT IS USEFUL**

The Village Database is an online compilation of county directories from Guangdong Province in China that the American Consulate in Hong Kong published during the 1960s. The database is based on four books, each book containing entries for a district (or county) in Guangdong province, where the vast majority of Chinese emigrants to North America originated from.

---

**THE TOISHAN INDEX OF CLAN NAMES** is one of four books that document residents and villages in Guangdong Province that are searchable on the Village Database. This is a sample page from the index.

<table>
<thead>
<tr>
<th>Cheung Shan Heung</th>
<th>7022-1472-8763</th>
<th>長山輝</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1st Area, Toishan</strong></td>
<td><strong>Markets:</strong> Toishan City 0669-1004</td>
<td><strong>Map Location:</strong> FQ 8269</td>
</tr>
<tr>
<td></td>
<td>Sai Moon 6007-7024</td>
<td><strong>Remarks:</strong> Cheung Shan Heung consists of three heungs:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1) Pak Hang Heung 0554-0677-8763 北坑輝</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2) Kwan Hou Heung 5028-0624-8763 騰厚輝</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3) Pang Sar Hang Heung 1756-3087-8766 彭沙坑輝</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Listed Villages or Lays:</strong> Pak Hang Heung</td>
</tr>
<tr>
<td>Pak On</td>
<td>6007-1344</td>
<td>7806 西安</td>
</tr>
<tr>
<td>Pak Hang</td>
<td>0554-0677</td>
<td>北坑</td>
</tr>
<tr>
<td>Wing On</td>
<td>3057-1344</td>
<td>永安</td>
</tr>
<tr>
<td>Cheung On</td>
<td>7022-1344</td>
<td>長安</td>
</tr>
<tr>
<td>Tung Shing</td>
<td>2639-4141</td>
<td>東盛</td>
</tr>
<tr>
<td>Nam Sing</td>
<td>0539-0524</td>
<td>南勝</td>
</tr>
<tr>
<td>Kau</td>
<td>5283</td>
<td>蕭</td>
</tr>
<tr>
<td>Miu Tsui Luk</td>
<td>1680-0731</td>
<td>7120 廟咀</td>
</tr>
<tr>
<td>Nam Cheong</td>
<td>0589-2490</td>
<td>7806 南昌</td>
</tr>
<tr>
<td>Tung Lok</td>
<td>0681-2887</td>
<td>同樂</td>
</tr>
<tr>
<td>Sai Hung</td>
<td>6007-0077</td>
<td>西亭</td>
</tr>
<tr>
<td>Sai Yuen</td>
<td>6007-0954</td>
<td>西園</td>
</tr>
<tr>
<td>Tung Sing</td>
<td>2639-2052</td>
<td>京威</td>
</tr>
<tr>
<td>Ko Lung</td>
<td>7559-7893</td>
<td>高龍</td>
</tr>
</tbody>
</table>

© 2021 Oregon Historical Society
THIS MAP of Guangdong Province, China, highlights the Sanyi and Siyi districts from where a majority of Chinese in Oregon migrated between 1850 and 1942.

Historian Him Mark Lai rediscovered the original four books, and the Village Database is the result of his desire to make a digital version available. The district books are invaluable primary sources for finding family lineages of Chinese descendants from villages in the Siyi region and Zhongshan County. Traditionally, Chinese genealogy lineages include only the names of males, usually by their marriage name. Only the clan surname of the wife is included. In addition to using the district books for investigating Chinese people seeking to fraudulently enter the United States as citizens or children of U.S. citizens — commonly known as paper sons or paper daughters — the U.S. consulate also used them for verifying and processing a large number of visas, passports, and federal benefit applications. There are approximately 14,000 villages and hamlets (small villages) covered by the four books; however, they do not include all of the villages found in the four counties in Guangdong Province. Each village name is listed with the primary associated clan surname (and occasionally more than one surname), the name of the county, area numbers of county heungs (sub-divisions), further subdivided by subheungs (sub-districts), and the names of the nearest market towns. The heung is keyed to the county map associated with each book and according to the grid coordinates used by the U.S. Army Map Service.

The original books for the Village Database list each village in three ways: by Standard Telegraphic Code (STC) number, romanized Cantonese name, and traditional Chinese characters. In the initial computer implementation of these four books as the Village Database, the STC numbers and simplified Chinese characters and Pinyin (the official Chinese romanization) were not implemented. The STC is a four-digit decimal code used in electrical telegraphing messages for representing Chinese characters — each STC number specifies a Chinese character. The romanization of Cantonese names uses Western words that are language-dialect specific to phonetically simulate the Chinese pronunciation of Chinese characters. The Village Database likely used Cantonese romanization because almost all the Chinese who came to America between 1849 and 1965 originated from Guangdong Province, where Cantonese was the lingua franca. The Cantonese romanization and traditional Chinese characters used in the Village Database are also used by the people of Hong Kong and Taiwan, and by many of the 50 million overseas Chinese throughout the world.

Many of the names of people, places, and villages found in documents stored at the National Archives and Records Administration (NARA) do not quite match with village names in the Village Database simply because...
The Village Database can help researchers confirm the name of their ancestors’ villages as well as clan names associated with the villages and their associated locations.

The qiaoxiang (ancestral village or home village) is the starting point for researching any member of the Chinese diaspora, including many Chinese Americans. Multiple villages associated with the same clan surname or village name can be differentiated in the Village Database by their locations within heungs, subheungs, or by their proximity to the nearest larger market town. Descendants of immigrants who want to see where their ancestors lived or learn about the old ways of village life may find that their ancestral village has a genealogy book that contains family or clan lineage and provides avenues for additional research.

The Village Database utilizes a hierarchical administrative system to organize locations documented in genealogy books. This system is designed to help researchers identify and track the lineage of their ancestors across generations and geographic regions.

**VILLAGE DATABASE HIERARCHICAL ADMINISTRATIVE SUBDIVISIONS**

- **County:** Toisan, Sunwei, Hoiping, Chungshan
- **Area:** County subdivisions numbered 1 to 10
- **Heung:** Name of heung or subdivision of a township
- **Subheung:** Name of subheung name
- **Village:** Name of village / village alias
- **Hamlet:** Name of hamlet (small village)

Since the establishment of the People’s Republic in China 1949, some village names have changed, and with the increased usage of simplified Chinese characters, simplified Chinese names for villages sometimes differ from the names originally written in traditional Chinese characters. Fortunately, original village names in NARA-archived documents, on Chinese gravestones, or in family documents are the same names in traditional Chinese characters recorded in the Village Database. Each village was designated with a village chief who is responsible for the affairs of the village and who visitors should contact with questions about village genealogies or when researching individuals or families.

While village clusters can usually be found on modern, digital maps, most of the smaller villages and hamlets may not appear there — but this does not mean they have disap-
peared. When using online mapping websites, and depending on the zooming capability and resolution of its digital cartographic database at the largest scale, some smaller villages and hamlets may appear. In cases where the smaller villages and hamlets are not displayed, researchers should consult Chinese gazetteers that record administrative changes to confirm whether villages still exist as distinct geographic entities or if they were consolidated within a larger village cluster.

SEARCHING AND BROWSING THE VILLAGE DATABASE

The Village Database main page provides two options for finding villages and surnames: search and browse. Under the search feature, users can search by clan surnames or by village names by their Cantonese romanization, traditional Chinese characters, and as of July 2020, by their Pinyin romanization. Searching by simplified Chinese characters and STC code are not supported. In the search feature, village names are displayed only in traditional Chinese characters and romanized Cantonese. Search results typically include one or more surnames associated with a village; however, there is usually just one clan surname for a village. These surnames are shown in parenthesis next to the village name. The screen capture on the facing page is a detail of a search for villages associated with the clan name Wong, which displays a map with heung locations with villages associated with the clan name. Using the browse feature, researchers can display a listing of villages by county, area number for the county subdivision, heungs, and subheungs. On the following pages, screen-captures of the search and browse features illustrate results for searches associated with Der Bin (謝邊 Xie Bian) village in Taishan County.

VILLAGE DATABASE MAP LOCATIONS

For each of the four Guangdong Province index books that compose the Village Database, the American Consulate General in Hong Kong created accompanying district maps. The maps indicate the location of heungs in the district using the Military Grid Reference System (MGRS), based on the Universal Transverse Mercator (UTM) grid system, and are documented in the Village Database. For example, Der Bin is located in Sam Sher (三社) heung, and its map location is FQ7869: FQ is the 100,000-meter square identifier, “78” is the Easting grid line value, which is an approximated value between the 70,000 and 80,000 grid lines, and “69” is the approximated Northing grid line value between the 60,000 and 70,000 grid lines. Although Der Bin’s FQ7869 marker indicates the approximate location of the heung, the district index to search by surname, users select from the pull-down menu that displays the romanized surname in Cantonese with the traditional Chinese character next to it. After selecting the surname, the database displays a regional map showing heung locations containing villages with that surname. It also lists all the villages associated with that surname, according to all the districts that contain a village with that surname.
THE VILLAGE DATABASE allows users to search by Cantonese romanization (top), traditional Chinese characters (middle), and beginning in July 2020, in Pinyin romanization (bottom). Note the search results include Toishan as the county, the area number 4, Sam Sher Heung, along with the village name.

THE BROWSE FEATURE displays listings of villages by county, area number for the county subdivision, heungs, and subheungs. The group clusters within a particular heung or subheungs of villages may include the village. Above is the browse result for Toishan County, Area 4, and Sam Sher Heung, with Der Bin village appearing as number seven in the Ha Sher subheung. The village details include Chinese characters (謝邊), Cantonese romanized name (Der Bin), and the associated clan surname (黃/Wong).
base maps do not provide enough detail to display the village names and their locations on a map. These MGRS locations have been available on the Village Database since 2001, but the built-in district mapping capability was not incorporated until mid 2020. Prior to the upgrade, I provided scanned copies of the original maps, held in the Geography and Map Division of the Library of Congress, to the Chinese Cultural Center of San Francisco, which is the organization that now owns the Village Database and the SiYi Forum, an international forum for Chinese genealogy research on the Siyi region of Guangdong Province. Researchers were able to access these maps as references when researching Chinese ancestors.

During the fall of 2009, the online Village Database was not available for several months. Since it was such a valuable tool for researching Chinese villages in the Siyi region, several members of the SiYi Forum, including me, wanted to set up an online clone of the Village Database. One member had already downloaded the Village Database data, and I had geo-referenced maps associated with the database to determine the approximate geographical coordinates of heungs. We halted our project when the Village Database came back up online. The charts I created with annotated grid lines, however, have been valuable in presentations at Chinese genealogy workshops that I organized across the United States — they provided important geographical coordinates for researchers before the recent inclusion of mapping in the Village Database. An example of one of the grid coordinate charts for Taishan is shown on page 566. In addition to using the built-in mapping features on the Village Database, researchers can also use online mapping tools, such as Google Maps, to transpose the general location of a heung, which could help identify the general area where villages are located.

SEARCHING FOR VILLAGES ON ONLINE MAP WEBSITES

Researchers should keep in mind that online map websites, whether in English or Chinese, have different search parameters and will yield different results. Searching online map websites for a village name by a romanized village name, in either Cantonese or in Pinyin romanization, does not guarantee a successful match. Searching a village name in traditional or simplified Chinese characters also does not always provide a successful match. Online map websites may show the general area where a village is located when searching for the names of heungs, subheungs, and market towns, but a visual search of this area on the map may be required. As previously mentioned, in many situations, villages have been absorbed by a heung or subheung and the consolidated villages and hamlets then assumed that name.

To find a village location, researchers may need to collectively confirm any or all of the following: the county

THIS IS AN EXAMPLE of a district index map that accompanies the four indexes of clan names. The maps indicate the location of heungs in the district using the Military Grid Reference System (MGRS), based on the Universal Transverse Mercator (UTM) grid system, and are documented in the Village Database. The maps, however, do not document places at the village level.
**This Grid** shows how to map geographic coordinates for Sam Sher heung in Taishan County. Before geographical coordinates were included in the online Village Database, these charts provided important location information for researchers. The bold numbers represent the UTM grid coordinates that are next to sexagesimal degrees.

In which it is located, heung, subheung, the names of nearby villages, and the nearest market town. Knowing the general area where the village should be, according the heung location as indicated in the Village Database, usually saves quite a bit of frustration during a process that can be long and arduous. For example, there are 1,126 villages associated with the clan name Wong in the Village Database, of which 678 are located in Taishan County. Searching for a Wong village in Taishan County by a person’s surname, such as Lung On (龍安), yields nine Wong villages associated with this name; two of them are located in the same heung and near the same market town. The only distinguishing features are the villages’ subheungs, which will be essential in locating them on a map.

Once the correct village is located on Google Maps, researchers may also want to determine the geographic coordinates of the village in order to visit in person. Locating the village on a Chinese map website, such as Baidu Maps, and making a copy of the map usually helps Chinese drivers locate the village. To obtain the geographic coordinates expressed in decimal degrees on Google Maps, right click a point on the map and the coordinates will be at the top of the drop-down list. Decimal degrees facilitate mathematical calculations directly, whereas geographic coordinates expressed as sexagesimal degrees, minutes, and seconds require conversion to decimal degrees for mathematical computations. The geographic coordinates for Der Bin village are expressed as: decimal degrees (22°30’79” latitude, 112°36’29” longitude); sexagesimal degrees (22°18’28.7”N 112°44’11.7”E).

**Finding Accurate Village Locations** on online mapping, such as Google Maps, is sometimes difficult because the only locational information provided in the Village Database is for heungs. When Der Bin’s location (FQ 7869 on the Taishan county map) is entered in Google Maps, it is about two miles, as the crow flies, from the actual location. With this information known, it is a matter of searching for the village name in the area. As previously noted, a village or
hamlet sometimes cannot be located on a map because the village name was changed or the village was incorporated into a larger village cluster. In such cases, local Chinese gazetteers can help clarify locations. Researchers need to be Chinese-literate, however, to be successful in such searches.

OTHER TECHNIQUES FOR RESEARCHING CHINESE VILLAGES AND GENEALOGIES

The Village Database covers information on villages for 90 percent of Chinese living in North America as well as for Chinese within many other countries, whose ancestors immigrated from the Siyi region and Zhongshan (Chungshan) until the mid-1960s. Researchers should know, however, that discovering the name and location of ancestral villages from the Village Database is just the beginning in their quest to recover their family history and Chinese ancestry. When researching ancestors from other regions of China, similar techniques apply as Village Database searches — finding villages first by province, then county, then heung and subheung — and village genealogy books help with that search. Although Red Guards destroyed many in China, thousands of Chinese genealogy books outside of China escaped destruction, including genealogy books for Northern China. Once the ancestral village is found, contacting the village chief in China can confirm the associated surname for the village and existence of a village genealogy book. Researchers looking at Guangdong province or other provinces in China can also try to locate nearby villages with the same associated clan surname. Such villages may have village genealogy books that could include part or all of the family lineage being researched — uncles, brothers, and first and second cousins may have resided nearby. For example, people living in a village named “Dragon” that no longer had enough land to support more houses might buy a nearby piece of land to establish a “New Dragon” village; the original “Dragon” village may then be called “Old Dragon” village. Alternatively, if a village became available, clans with a different surname might move there and rename it. This is likely how some villages came to have more than one associated surname.

Researchers can also try to identify and locate nearby ancestral halls or temples with their ancestor’s clan surname to inquire about whether they hold lineage tablets or a clan genealogy book. The lineages in such a clan genealogy book may not go down to the most recent generations, as a village genealogy book would, but if researchers have family lineages documenting back three or four consecutive generations (traced through direct father-son descent), they may be able to find common ancestors. Then, the family lineage can be linked into clan lineages. Some knowledge of Chinese genealogy is needed to navigate this last technique as the only known information available is a man’s Chinese name and where he was born. In the Siyi region, most men have two names: a milk (birth name) and a hao (marriage name). The milk name is a name of expectation, reflecting some qualities that parents hope their son can attain. The hao is given when a man marries, and men are listed in the village genealogy book according to their hao name, which is composed of three Chinese characters. The first character is the surname, which is announced in the beginning of the genealogy book and not repeated for each name; it is understood that the Chinese character for the surname precedes all the two-character names in the lineages of the genealogy book. The second character in the name is usually the bon pie
If your life is prosperous and flourishing, even your friends will continue furnishing or establishing temples in remembrance of you and your family. This helps to establish a cultured and civilized society.

If your family continues to be successful generation by generation, it continually benefits the world and makes it more beautiful. Its success is secured only while you are an honest, truthful, and virtuous person with a moral life.

Doing kindly deeds is very helpful for your family and country. If you are successful at this and study it well, it may help you to be remembered for generations.

This Tom generational poem was translated by Yongping Zhao and edited by Henry Tom.
In 2007, when researching the family lineage of a Tan clansman, I asked my researcher in China to call the Xinhui County government and ask how many villages in that county the Xinhui County government and asked my researcher in China to call the county's genealogy book, however, had been destroyed. When I checked an associated clan genealogy book, I found that man's name and sent the family lineage to his great-grandson. The great-grandson was rather skeptical until he saw an adjoining page that had a photograph of an uncle, who was also included in the family lineage. On a subsequent visit, the great-grandson brought a copy of the genealogy back to his ancestral village. Stories such as these are incredibly rewarding and reinforce the importance of this time-consuming work.

Since 2005, I have served as a moderator, with the handle name Location, Location, Location, in the SiYi Forum, and I have personally assisted many overseas Chinese to start researching their family history. Prior to the beginning of the COVID-19 pandemic, I organized three-day Chinese genealogy workshops twice a year in San Francisco, California, and Las Vegas, Nevada. These workshops have helped Chinese Americans to understand and appreciate their Chinese ancestry and family history. In 2018 and 2019, I also organized workshops in China, where researchers visited historical sites of importance for Chinese Americans in Guangdong province, were immersed in rural Chinese village life in the Siyi region, and then visited their ancestral villages. As long as I am able, I will try to keep on organizing Chinese genealogy workshops to continue this important work of documenting Chinese American history.

NOTES

1. The author confirmed these percentages in discussions with historian Him Mark Lai at “Branching Out the Banyan Tree: 2005 Chinese American Studies Conference,” presented by the Chinese Historical Society of America with Asian American Studies Department at San Francisco State University, October 6–9, 2005. Usage of romanization of Chinese varied considerably prior to the use of Pinyin, the official standard romanization. These variations are reflected in this article according to its use in different publications. Pinyin is parenthesized after various romanized Cantonese names.

2. Him Mark Lai was a world-renowned historian, writer, and Chinese community activist. For more on his work and associated digital archives, see https://himmarklai.org/(accessed September 20, 2021). The Village Database is currently maintained by the Friends of Roots and can be accessed at https://villagedb.friendsofroots.org/search.cgi (accessed September 20, 2021).

3. The original four books contain information on approximately 14,000 villages and hamlets, but with the recent addition of villages, the number should be updated to 1,706 Enping villages, which would be approximately 15,700 villages total.


7. Researchers should note that the Easting and Northing numbers for the grid coordinates for heungs in Hoiping county were reversed.

HENRY TOM led a research group to China in November 2019, and they are pictured here at the entrance to Dr. Sun Yat-sen residence and museum in Zhongshan. Tom is kneeling in the center front, seventh person from the left.